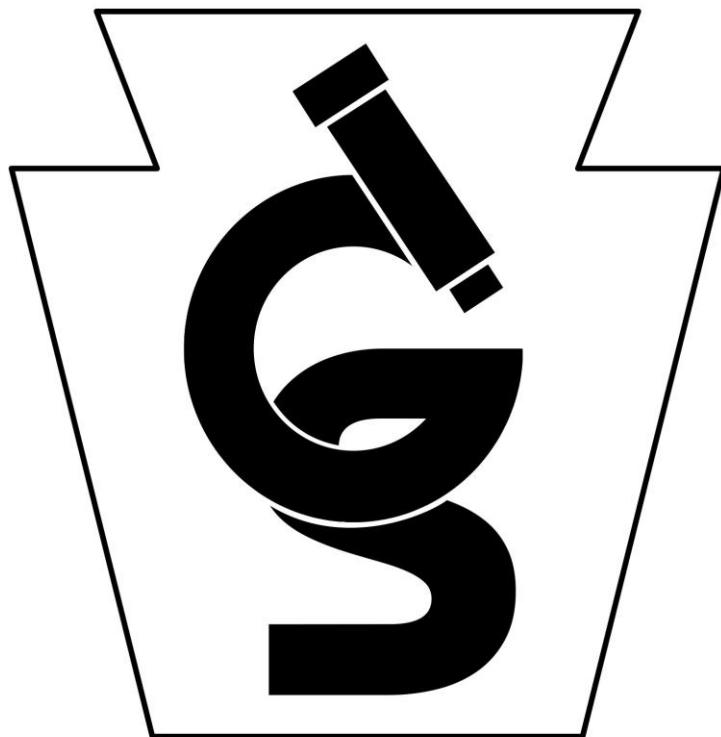


**Journal of the  
Pennsylvania Governor's School for the Sciences**

Class of 2023



Volume 38, 2023

Mellon College of Science  
Carnegie Mellon University, Pittsburgh, PA 15213



Journal  
of the  
Pennsylvania Governor's School for the Sciences

*Class of 2023*

**Volume 38, 2023**

**Copyright (c) 2023, by The Pennsylvania Governor's School for the Sciences  
and Carnegie Mellon University**

**Permission** is granted to quote from this journal with the customary acknowledgement of the source. To reprint a figure, table or other excerpt, requires, in addition, the consent of one of the original authors and notification of the PGSS.



**The Pennsylvania Governor's School for the Sciences**

In cooperation with



**Major funding provided by**



**(Alumni, Parents and Friends of PGSS)**

**Corporate Partners:**



## **Educational Partners:**

Pittsburgh Public Schools  
Lancaster-Lebanon Intermediate Unit (IU 13)  
Selinsgrove Area School District  
Central Susquehanna Intermediate Unit (IU 16)  
Loyalsock Township School District  
Chester County Intermediate Unit (IU 24)  
The School District of Philadelphia  
CCRES, Inc.

The PGSS program and its director  
gratefully acknowledge the support of  
Carnegie Mellon University and the  
**Mellon College of Science.**

**Carnegie Mellon University**

## Table of Contents

<b>Preface .....</b>	<b>1</b>
<b>Acknowledgements .....</b>	<b>2</b>

### **Biology Team Projects**

<b>DNA Damage &amp; Ultraviolet Radiation: Efficacy of Various SPF Containing Products on <i>Saccharomyces cerevisiae</i></b>	
Barlett, Cudzil, DeAlmeida, Dumasia, Feng, Griffith, Ramos, Hu, Huang, Liu, O'Connor, Tesfaye, Wong .....	<b>5</b>
<b>Investigating the Impacts of Environmental Factors on Neuron Function Using <i>L. terrestris</i> and <i>C. elegans</i></b>	
Aguay, Balaji, Breithaupt, R DeAlmeida, Jindal, Karunaratne, Nair, Stackhouse, D Thomas, Zhang .....	<b>45</b>

### **Biological Physics Team Project**

<b>Determining the Secondary Structure of a Cyclic Antimicrobial Peptide Using Circular Dichroism</b>	
Rogers, Suryawanshi, Umesh, Viswanathan, Xie, Xu .....	<b>77</b>

### **Chemistry Team Projects**

<b>Analysis of Atmospheric Trends and Climate Change Impacts using Pittsburgh Weather Balloon Data</b>	
Ram, Ravi, Sajwan, Saldubehere .....	<b>99</b>
<b>A Breath of Fresh Air? Analysis of Outdoor and Indoor Air Quality and Exposure in Pittsburgh</b>	
Beeson, Ehrensberger, Peterson, Vasbinder .....	<b>125</b>

### **Computer Science/Math Team Projects**

<b>Journey Through the Star: A Minimax Exploration of Chinese Checkers</b>	
Busche, Y Liu, Mao, Mukherjee, Nguyen, Parikh .....	<b>163</b>
<b>Connecting the Dots: Using Turn-Based Game Strategy to Solve Dots and Boxes</b>	
Abraham, Barsotti, Beckish, Chan-van der Helm, Y Kim, Rajesh .....	<b>191</b>
<b>Paper Soccer: Exploring and Evaluating Different Algorithmic Approaches Toward Optimal Play</b>	
Bhat, Mei, Singh, D Wang, E Xu.....	<b>219</b>

**Electrical Engineering Team Projects**

**Computational Modeling and Simulation of Inductive Loop Detectors for Vehicle Detection, Classification and Motional Analysis**

Philip, Tharumia, Tomy, F Yang ..... 245

**Metal Detection Techniques**

Grolemund, Hou, Kong..... 271

**Physics Team Projects**

**Long ago, in a Globular Cluster Far, Far Away**

Gurreonero, Hasty, Ranaweera, Seidman, C Thomas ..... 291

**Light: Wave, Particle or Something Else?**

S Lin, Anokye ..... 325

**Crystal Structure and Differences in Magnetic Behavior in Copper Sulfate and its Hydrates**

Brown, Risch, R Wang ..... 339

**Appendix**

The Students, Faculty, and Teaching Assistants/Counselors

of the Pennsylvania Governor's School for the Sciences for 2023 ..... 355

## Preface

The Pennsylvania Governor's School for the Sciences (PGSS) is a five-week summer residential program for talented Pennsylvania high school students. The PGSS was initiated by the Pennsylvania Department of Education (PDE) in 1982 and is hosted by the Mellon College of Science on the Pittsburgh campus of Carnegie Mellon University.

The PGSS class of 2023 consisted of 71 students – 35 male and 36 female – chosen from a pool of 403 applicants from all over the Commonwealth of Pennsylvania. The 38th session of PGSS was conducted in 2023 from July 2 through August 5.

The PGSS academic program consists of five key components:

- Core Courses in biology, chemistry, computer science, mathematics and physics provide a common educational experience for all students.
- Elective Courses, which vary from year to year, may be chosen according to the interest of each individual student.
- Formal Laboratory Courses are offered in biology, chemistry, computer science, mathematics, and physics. Each student chooses one formal laboratory course.
- Panel discussions on topics of current interest are required of all students.
- Team Research Projects are offered on a wide variety of topics. Each student chooses one project and works with a team of students under the supervision of a faculty member.

The PGSS Team Research Projects involve the investigation of an original problem or the solution of a problem by techniques original to the student investigators. While a faculty member is available for initial direction and advice, most of the accomplishments come from the students' own initiative.

The Journal of the Pennsylvania Governor's School for the Sciences reports the results of the students' efforts and is the official record of these team research projects. Each team investigated their chosen problem using resources and techniques appropriate for that topic. The following reports were written by the student team members. The faculty advisor for each project reviewed the final report and made necessary minor corrections.

This journal was edited by Melissa Lessure, Assistant to the Director of PGSS, who assumed the responsibility of converting all the reports to a similar format, reproducing figures, and dealing with other publication issues. Otherwise, the original character of the student-authored papers was maintained to the maximum feasible extent.

## **Acknowledgements**

The Pennsylvania Governor's School for the Sciences was supported this year by grants from the PGSS Campaign, Inc. (Alumni, Parents and Friends of PGSS) and the Pennsylvania Department of Education. We are particularly grateful this year for donations from our corporate sponsors: Air Products and EQT Foundation. The PGSS gratefully acknowledges the continuing support of the Mellon College of Science at Carnegie Mellon University and generous contributions from the parents and guardians of the PGSS Class of 2023.

Barry B. Luokkala, Ph.D.  
Program Director of the Pennsylvania Governor's School for the Sciences  
October 2023

**BIOLOGY  
TEAM PROJECTS**



# **DNA Damage & Ultraviolet Radiation: Efficacy of Various SPF Containing Products on *Saccharomyces cerevisiae***

Grace E. Barlett, Hannah Grace Cudzil, Dinuk R. DeAlmeida, Suraj Rupal Dumasia, Harry P. Feng, Cassidy Griffith, Marbella Ramos Guardado, Zheyu Hu, Viki Huang, Dorothy Liu, Kaila O'Connor, Alem S. Tesfaye, Meghan C. Wong

## **Abstract**

Ultraviolet light can induce DNA mutations, which are repairable to an extent. However, with constant UV radiation exposure and without SPF protection, DNA damage can accumulate over time and overwhelm the DNA repair mechanisms. When enough DNA damage occurs without repair, maladies like skin cancer, premature aging, and sunburn can arise. To study how this occurs in human cells, our project investigates protective measures against UV radiation on a model organism, *Saccharomyces cerevisiae*. The cells were exposed to radiation at various time points while implementing protection from a wide range of SPF-containing products. We then examined cell viability rates and DNA damage that resulted from UV radiation. Specifically, we investigated cell viability using killing curves, nuclear fragmentation, and indirect immunofluorescence assays. By examining this data, we found that exposure to UV radiation induces mutations and DNA damage, ultimately leading to apoptosis in *S. cerevisiae*. From the gel electrophoresis results, we observed fragmentation of DNA exposed to radiation, allowing the visualization of the damage inflicted on DNA over several time points. The immunofluorescence process implemented antibody recognition of DNA damage-induced phosphorylation to visualize the damage to yeast cell nuclei. Viability assays of several SPF products allowed us to compare the efficacy of different sunscreen types (spray vs lotion), and to determine the efficacy of water-resistant, antioxidant, and expired sunscreens in protecting the cells from UV radiation.

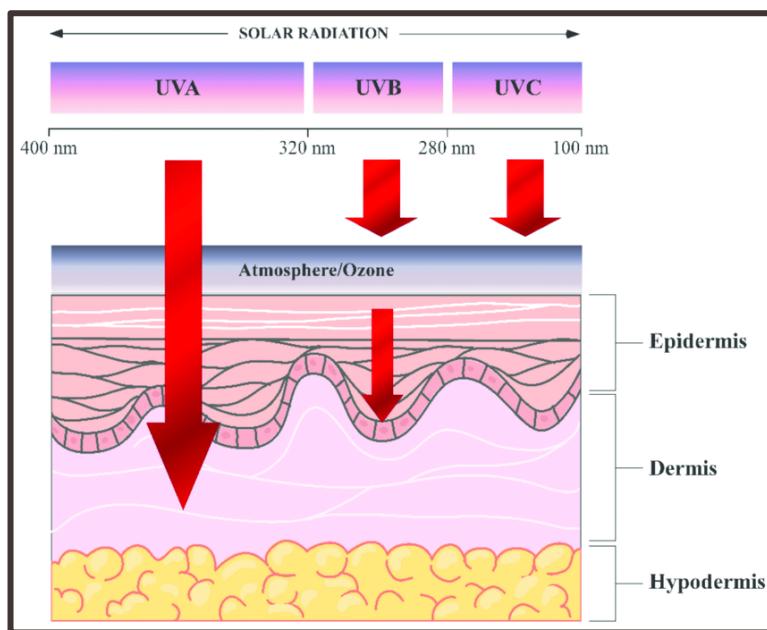
## **I. Introduction**

### **A. Skin Cancer in the United States**

Cancer continues to be one of the most common causes of death in the United States, having a mortality rate of around 1,670 people each day in 2023<sup>1</sup>. Cancer affecting the skin is the most common form in the United States. It is estimated that 1 in 5 Americans will develop skin cancer in their lifetime<sup>2</sup>. The prevalence of skin cancer-related deaths is continuing to increase; for example, the total incidence rate of melanoma doubled between 1982 and 2011 and is projected to increase in the next 15 years<sup>3</sup>. The mortality rate has risen to 4.28%, with 186,680 total cases and 7,990 deaths documented from 2022 to 2023<sup>4</sup>. Reasons for this increase may have to do with global warming, the subsequent increase in UV radiation levels, as well as our increased ability to identify skin cancer as the cause of death. Regardless of how these reasons overlap, it is essential to be proactive about this issue with preventative measures that will be discussed later in this paper.

## B. UV Radiation

Ultraviolet (UV) radiation is the most common cause of several skin cancers, including melanoma<sup>5</sup>. UV radiation is a type of invisible electromagnetic radiation, consisting of photons traveling at the speed of light and in a wave-like pattern. UV radiation is strongest during the hours of 10 AM and 4 PM when the sun has the shortest path to the Earth<sup>6,7</sup>. Other factors can affect the amount of UV that reaches the earth. UV rays are usually stronger in the spring and summer when they concentrate at the equator. Additionally, UV rays are more likely to reach higher elevations and less likely to penetrate cloud cover.



**Figure 1: Types of UV Radiation and Skin Penetration<sup>8</sup>**

Figure 1 shows the spectrum of ultraviolet radiation and how each form of UV Radiation interacts with layers of human skin.

The different types of UV radiation are UVA, UVB, and UVC rays;<sup>9</sup> each is characterized by wavelength and how it penetrates the human skin (see Figure 1). UVC rays are 100-279 nm and they emit the most energy out of all the UV rays.<sup>11</sup> Exposure to possible UVC rays could lead to severe burns on the skin and eyes. These rays are completely absorbed by the atmosphere and ozone layer.<sup>9</sup> UVB rays occur at 280-314 nm. While most are absorbed by the ozone, some are still able to reach the earth's surface.<sup>11</sup> UVB penetrates the epidermis and is specifically absorbed by cellular DNA. These rays contribute to skin aging, sunburns, and the development of skin cancer.<sup>10</sup> UVA rays are typically 315-400 nm and contain the least energy.<sup>11</sup> In addition, UVA rays are not absorbed by the ozone layer at all. While UVA is not directly absorbed by DNA like UVB, its rays can still penetrate the skin and cause indirect DNA damage.<sup>1</sup>

Degradation of the stratospheric ozone layer is a major factor for increased UV exposure and higher skin cancer mortality rates. The ozone layer is a barrier that forms a protective layer 6-10 km above the Earth, filtering out UVC and most UVB<sup>12</sup>. From the 1970s to 2019, over 70% of the ozone concentration over Antarctica degraded, creating an “ozone hole” over the continental region<sup>10</sup>. Ozone destruction had been catalyzed by chlorine and bromine compounds in pollutants such as chlorofluorocarbons, organo-bromides, methyl chloroform, etc., eventually leading to a 10 million square mile hole in the ozone layer<sup>13,14</sup>. Currently,

the ozone hole has shrunk for the past fourteen years due to new global policies such as the Montreal Protocol, which regulates the production of the aforementioned ozone-depleting substances<sup>15</sup>. However, this does not fully mitigate the problem of UV radiation. The repairing of the ozone layer will still take a significant amount of time and is affected by many factors such as global warming and the remaining ozone-depleting substances in the atmosphere. Most importantly, some UV rays can still penetrate the ozone and affect humans, further enforcing the prevalence of this issue.

### C. Health Effects of UV Radiation

Because UV radiation is always present and capable of penetrating the skin, it is essential to understand the health effects of UV radiation. For example, prolonged exposure to UV rays may lead to erythema and dermatitis, commonly referred to as sunburn. This event leads to an increased risk of melanoma, BCC, and SCC because it provides for an accumulative amount of DNA damage that the body will have to remedy<sup>16</sup>. Photoaging—degradation of the collagen on the skin's surface because of a lack of UV protection—also exemplifies one of the seemingly minor effects of UV radiation<sup>17</sup>. Individuals at significantly young ages will exhibit skin conditions such as wrinkling and chronic dryness because of UV exposure, showing the necessity of UV awareness<sup>16</sup>.

Spending long days at the beach and receiving tans may be attractive ideas, considering that they build on the aesthetics of changing skin pigmentation through the sun's rays and overall represent enjoyable activities. However, they also reveal a prominent yet insidious effect of UV rays because they cause deleterious effects invisible to the naked eye. Furthermore, UV rays do not just lead to aesthetic changes; they are also correlated to several acute effects on the skin. These include hyperkeratosis, which is the thickening of the skin due to an abundance of cell repair and reproduction. The accumulation of minor effects like hyperkeratosis can lead to debilitating, chronic conditions including multiple sclerosis, allergic asthma, and diabetes. Although these diseases seem very different from one another, they all involve the deficiency of vitamin D3. This shows how the increase of SPF protection and/or protection from melanin leads to less vitamin absorbance, which can have unexpected adverse effects<sup>18</sup>. In addition, as the skin becomes more susceptible to its environment due to increased UV exposure, the person may experience overall immunosuppression that extends to combating pathogens and resisting UV damage. As mentioned earlier, UV radiation aggravates keratinocytes to abnormally multiply and thicken the skin, but it can also make them produce immunosuppressive soluble mediators, leading to the suppression of inflammatory and repairing responses<sup>19</sup>. These weaknesses in the immune system can easily lead to photoaging, sunburns, and atrophy, showing how the negative effects of UV rays blend into a nuanced problem.

However, the effects of UV radiation can also be positive. Being in the sun tends to improve mood by releasing endorphins, which is the scientific basis for the idea of phototherapy. It also provides for the synthesis of Vitamin D within the body<sup>20</sup>. The relationship between melanin, vitamin D, and skin health is complex because melanin protects skin health but can simultaneously cause a vitamin D deficiency. Melanin is considered highly protective, but it also has its downfalls<sup>21</sup>. The production of too much melanin results in the creation of reactive oxygen species, which contributes to aging and inflammation of the skin. Because reactive oxygen species produce these effects, they may play a role in inflammatory skin conditions like contact dermatitis, urticaria, and atopic dermatitis<sup>22</sup>.

### D. Types of Skin Cancer

Basal cell carcinoma (BCC) is the most commonly diagnosed form of skin cancer. Each year, approximately 3.6 million new cases of BCC are diagnosed in the United States. BCC formation is initiated when UV rays penetrate the basal layer of the epidermis. The basal layer is responsible for generating new skin cells that cause mature skin cells to rise in the epidermis; by generating new cells, older ones rise through the various

levels of the dermis and epidermis. While UV radiation from both UVA and UVB contribute to the destruction of DNA in basal skin cells, it is believed that UVB radiation causes more damage due to UVB's shorter wavelengths. When many of the BCC cells are elevated in the epidermis, the BCC cells accumulate at the skin surface. If the BCC cells do not divide rapidly, the BCC can be detected and treated at an early stage of development. However, many individuals that are eventually diagnosed with BCC experience a 20-50 year latent period before BCC is manifested. In BCC patients, accounts of exposed sores, bumps, abnormal pink growth, red patches, and abnormal cellular growth are the hallmarks of UV damage to the skin, specifically to BCC. Other risk factors like tanning bed use, family history of skin cancer, skin type, immunosuppression, and previous exposure to radiotherapy can contribute to the unique approach to BCC patients' treatment. In the majority of BCC cases, localized skin cancer rarely metastasizes, but cancer can become locally invasive to the point where BCC can spread deeply to surrounding tissues and bone.

Squamous cell carcinoma (SCC) is considered the second most common form of skin cancer. In the United States, approximately 1.8 million new cases are diagnosed each year. Similar to the impact of UV rays in basal cell carcinoma, UV rays penetrate the squamous cell layer of the epidermis resulting in DNA damage. Specifically, the UV damage in squamous cells causes a mutation in the p53 tumor suppressor gene that leads to uncontrollable cell growth. Furthermore, SCC formations are most likely to be present in the middle and outer layers of the skin and in areas of the body that receive the most sun exposure (e.g. head, arms, and legs). Unlike BCC, SCC can be classified as either cutaneous or metastatic. In metastatic SCC, SCC can spread to other areas of the body, while cutaneous SCC affects only the top layer of the skin. Common hallmarks of SCC include rough-feeling bumpy skin, a depression in skin growth, a wound that does not heal, and an area of red, scaly skin that covers approximately one inch of the skin surface.

The Skin Cancer Foundation estimates that 7,990 people will die of melanoma in 2023<sup>4</sup>. Melanoma is classified as a cutaneous malignant growth caused by cells found in the deepest layers of the epidermis, containing melanocytes. The initial cause of melanoma is a mutation that occurs within the melanocytes often due to a disturbance that affects the cell growth and reproduction of the cell (mitosis) causing the formation of cancer. One of the most common causes of melanoma is from sun exposure as recent reports linked about 86% of melanoma cases to UV exposure<sup>23</sup>. Melanoma is usually found in places where UV exposure is most concentrated in parts of the skin. Due to the damaging effects of UV on melanocytes, it disrupts the basic function of these cells causing rapid, irregular growth. The rapidness of the growth can cause a higher severity when melanoma progresses to later stages. Melanoma is normally marked by unusual formations in the skin that were previously there or newly formed. Some unusual markings that may be associated with the formation of melanoma include moles, sores, blemishes, and lumps. Likewise, melanoma can occur in various parts of the body besides the skin. The mouth, eye, finger, and toenails all are other places that were reported to be possible melanoma sites<sup>24</sup>.

To understand the unusual formations of melanoma, one may use a skin guide, like the ABCDE rule from the Skin Cancer Foundation, to analyze signs and provide key tips for spotting melanoma<sup>25</sup>. The A in the rule stands for the asymmetry of the unusual marking, considering if the blemish or mole is asymmetrical or matching on both sides. The B in the rule stands for its border, as with most signs of melanoma, the blemish has unclear borders with irregular shapes. The C in the rule stands for the color, as melanoma is usually in different shades, such as brown or black, and sometimes with patches of other colors like white or red within the unusual skin formation. The D in the rule stands for diameter, unusual skin formations that are linked to melanoma start usually around 6 mm across in length, however, sometimes they can start smaller than this. Lastly, the E in the rule stands for evolving, notice if the unusual skin formation changes shape, size, or color over time. This is a sign of melanoma. Besides the uniqueness of the structure and appearance of melanoma within its first formations, there are more than just structural changes and physical attributes that contribute to the understanding of melanoma.

To understand what melanoma is further, consider the processes and methods behind measuring and tracking the unique spread of melanoma from the skin to the body. One method commonly used is staging<sup>26</sup> (Figure 2). A stage can take into account the amount of melanoma throughout the body. The stage's quality is essential in determining the severity of the cancer. Stages in staging are split up and can be characterized by numbers and letters. For example, the earliest stages of melanoma are considered stage 0 or stage 1, as most staging systems range from 1 through 4 stages. These 4 stages can be characterized even more by the use of capital letters and by a simple rule, the earlier the letter means the lower the stage. This simple construct of how the staging system works can be applied further.

Various staging systems exist, as some systems range in complexity and variation on how characterized the melanoma is, this helps pinpoint individual cases of melanoma and how to treat it directly. Consider the TNM staging system<sup>27</sup>, created by the American Joint Committee on Cancer (AJCC)<sup>28</sup>. The TNM system is based on 3 key factors in how it determines its stages in its system. The first factor can be linked to the first and second stages of melanoma, the factor is based on the tumor (T). The tumor is the first factor categorized by both tumor thickness and ulceration. Tumor thickness is the measurement of melanoma specifically called the Breslow measurement. The Breslow measurement takes into account the size of the growth of the primary tumor. For example, if the tumor's thickness is less than 1 mm then it has a smaller chance of spreading. On the other hand, ulceration, the breakdown of the skin over melanoma, acts as a benchmark for the first factor. The primary tumor continues to evolve to the point where it is around the equivalent of 4 mm in thickness. This point is where most ulceration occurs, as ulceration generally characterizes the second stage. The second factor that determines the stages is the spreading to the lymph nodes (N). This characterization occurs during the later stages such as stage 3. Scientists consider melanoma cancer cells that break away from the primary tumor; these cells enter the sentinel lymph nodes and then pass through the lymph vessels where they attach and form new tumors. They include the formation of various categories that melanoma metastasis is sectioned into (in-transit metastasis, satellite metastasis). These addressed forms of melanoma metastasis act as a determining characteristic for the later stages in melanoma staging. Lastly, the determining factor of metastasis (M) to distant sites. This condition is marked by the spread of melanoma metastasis throughout the body, as it can be ubiquitous it can spread to sites like the brain, bones, lymphatic system, liver, and lungs.

To classify an overall stage, there is another set of numbers and letters that provide more identification when paired with one of the three factors from the TNM staging system. For example, if there is a higher number after one of the factors (T, N, M) then that means that the cancer is more advanced. Once the T, N, and M factors have been determined they combine this information with another system called a stage-grouping process. Upon contribution, this completes a thorough process of determining and forming the stages in the staging process. This complex system however contributes to a lot of research in knowledge of prevention against melanoma and its survivability upon prognosis. Furthermore, this can be shown through melanoma's 5-year relativity survival rate.

Usually given 5 years, the survival rate is an analysis of the percentage of people with the same type of cancer that are still alive after they were diagnosed. The 5-year relative survival rate is linked with the staging process in tracking the spread<sup>23</sup>. The staging process creates a more accurate and precise method of characterizing melanoma in stages. Knowing the stages of melanoma contributes to the survival rate by giving the researchers an accurate estimate of what stages to compare the individuals at. The two key factors that are being compared and considered are both the stages of melanoma and the type of melanoma. It compares these two factors within the consensus of the individuals who have been diagnosed in the past five years. The higher the survival rate the higher the chance percentage that you will live the next five years. The relative survival rates rely on the SEER (Surveillance, Epidemiology, End Results program) database which tracks these rates specifically. However, it does not have the same methods as

the TNM staging method. Instead, the SEERs Database uses a different method. The first stage, the localization stage, is categorized as no sign that the cancer has spread. The second, called the regional stage, is when the melanoma has spread to the lymphatic system. Lastly, the third stage, which is labeled as the distant stage, is usually around the fourth stage of diagnosis; it's marked by the spread of melanoma throughout the body. All of these topics contribute to the understanding of what melanoma is and how it is treated as a cancer of the skin.

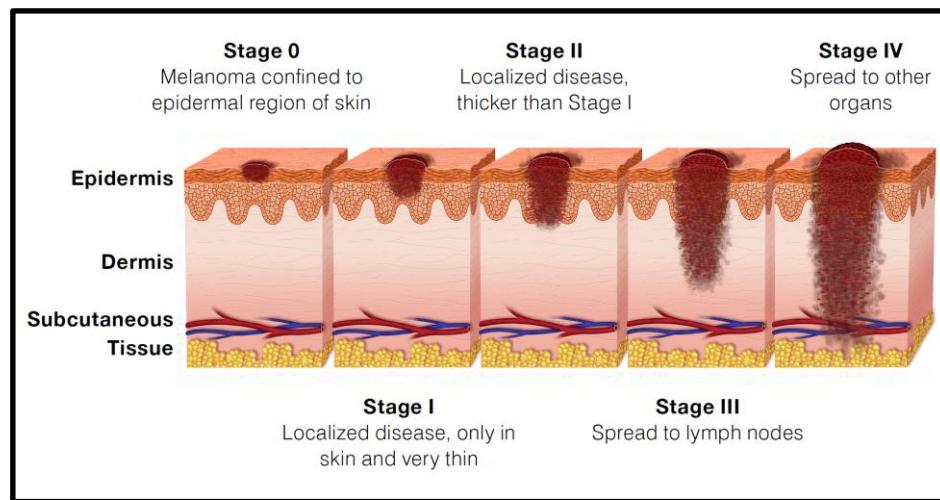
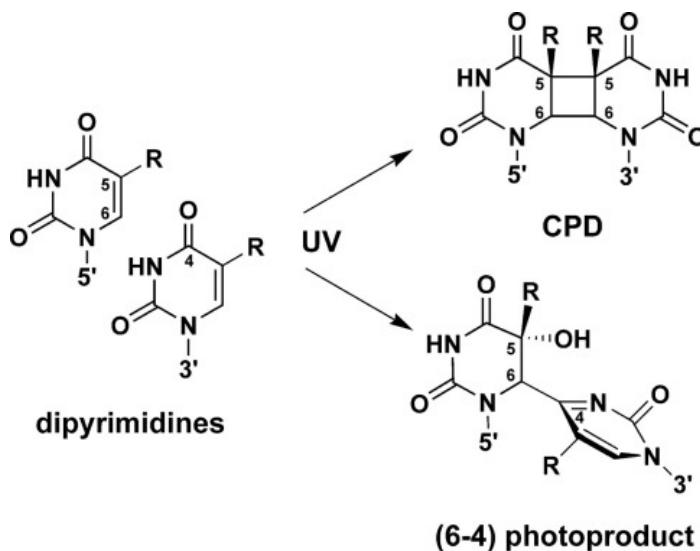


Figure 2: Cancer Staging of Melanoma<sup>29</sup>

Figure 2 displays the progression of melanoma through stages 0-4.

## E. DNA Damage

Ultraviolet radiation constantly strikes skin cells, which can damage the DNA in the nucleus. When ultraviolet radiation strikes two adjacent pyrimidines (thymine and thymine, cytosine and thymine, or cytosine and cytosine), these two pyrimidines can become excited, forming 2 single bonds between the adjacent bases (Figure 3)<sup>21</sup>. This bond between adjacent pyrimidines is called a cyclobutane-pyrimidine dimer (CPD). This dimer creates a bend of 7-9 degrees in the DNA<sup>10</sup>. Another type of DNA damage from UV radiation is 6-4 photoproducts (6-4 PP). When ultraviolet radiation strikes the DNA, two adjacent nucleotides are energized. A single bond forms between the nitrogen bases of these adjacent nucleotides in an intermediate state. Then, in one of the nucleotides, the bond between the nitrogen base and the deoxyribose breaks, forming a 6-4 PP. This lesion is more lethal than CPDs<sup>21</sup>. CPDs are responsible for about 75% of all DNA mutations after UV exposure, and 6-4 PPs account for about 25% of the DNA mutations from UV radiation. 6-4 PPs create a bend of about 44 degrees. These lesions in the DNA sequence can have many negative consequences for cells, such as inducing apoptosis. With CPDs and 6-4PPs, DNA polymerase (an enzyme that facilitates cell replication) is unable to read the DNA sequence. If these alterations are not fixed, then the cell cannot replicate DNA. Without replicating DNA, the cell cannot undergo mitosis. This inability to divide induces cell death. Additionally, RNA polymerase is the enzyme that facilitates transcription. However, RNA polymerase also cannot read DNA sequences with CPDs or 6-4PPs. This means that mRNA may not be produced from genes with CPDs or 6-4PPs. This stops all protein production that is related to the mutated genes, which can significantly impair the cellular functions that this protein is responsible for<sup>10</sup>.



**Figure 3: Formation of Cyclobutane Pyrimidine Dimers and 6-4 Photoproducts<sup>30</sup>**

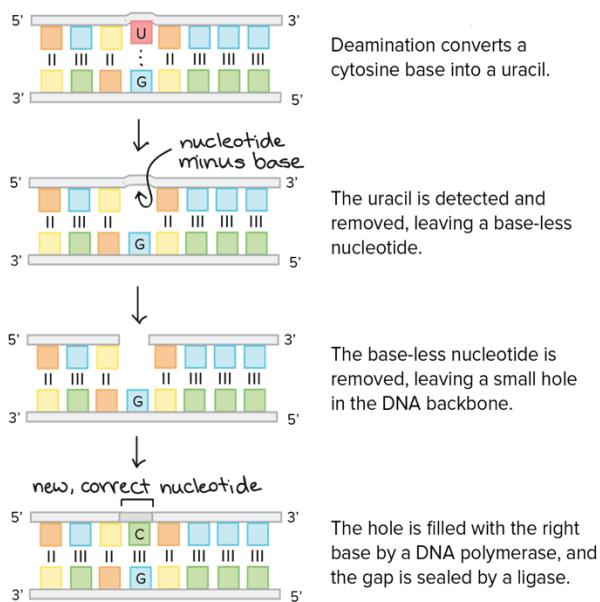
Figure 3 shows UV light-inducing DNA lesions by energizing two pyrimidine bases which result in either Cyclobutane Pyrimidine Dimers (CPD) or 6-4 photoproducts (6-4PP)

Some examples of genes that are affected by CPDs and 6-4PPs are PTCH1, EGFR, and BRAF. PTCH1 encodes a component of the sonic hedgehog signaling pathway and is involved with tumor suppression. Mutations in PTCH1 disrupt the signaling pathway, which can induce uncontrolled cell proliferation and cancer. Additionally, the PTCH1 gene is important to suppress tumors; therefore, a mutation in this gene can cause skin tumors to develop. Since disruption in the expression of the PTCH1 gene causes tumors to develop, the normal function of the PTCH1 gene is thought to be a tumor suppressor<sup>31</sup>. The EGFR gene is important to facilitate cell proliferation. Typically, four ligands must bind to the gene for cell proliferation to initiate; however, when UV causes mutations in EGFR, these ligands are no longer necessary to activate EGFR, causing continuous expression, cell proliferation, and uncontrolled cell growth. Since an increase in the expression of EGFR can cause cancer, it is classified as an oncogene<sup>32</sup>. Finally, the BRAF gene is a proto-oncogene that facilitates cell growth and division. The protein that is encoded by the BRAF gene helps to regulate the MAPK signaling pathway, which is involved in facilitating cell division. Specifically, the protein encoded by the BRAF gene can only stimulate the MAPK pathway if the RAS protein activates the protein that the BRAF gene codes for. Therefore, the RAS protein is necessary to initiate cell division. However, when UV damage causes a mutation in the BRAF gene, the protein encoded by this gene is misfolded. This misfolded protein continuously activates the MAPK pathway, regardless of activation by the RAS protein. This leads to uncontrolled cell division and skin cancer<sup>33</sup>.

UV light can also induce mutations in DNA through UV-signature mutations, as well as triplet mutations. UV signature mutation occurs when CPDs remain unrepaired, while triplet mutations are a specific type of genetic mutation that involves the simultaneous alteration of three consecutive nucleotides in the DNA sequence<sup>34</sup>. These mutations can result from various factors, including exposure to mutagens like UV light, ionizing radiation, or chemical agents<sup>35</sup>. When the DNA undergoes replication or repair, errors may occur during the process, leading to the mispairing, insertion, or deletion of three adjacent bases. Triplet mutations can have significant consequences, as they can disrupt the reading frame during translation, resulting in the incorporation of incorrect amino acids into the protein sequence. This can lead to the production of non-functional or aberrant proteins, affecting normal cellular functions and potentially contributing to various genetic disorders or diseases<sup>36,37</sup>.

## F. Excision Repair

The cell cycle is a series of events that takes place when a cell is getting ready and is in the process of dividing. The phases of the cell cycle include interphase, where the cell grows and replicates DNA, and mitosis, which is the division of the cell into two daughter cells. Within interphase, there are 3 stages: the G1 phase is where the cell prepares to divide and primes the cell for replication by making sure the cell has enough nutrients for division; the S phase is where the cell makes a copy of all DNA; the G2 phase is where cell prepares for mitosis by making sure the cell has sufficient nutrients to proceed through mitosis. The stages of mitosis include prophase, where the chromatin condenses and mitotic spindle forms, metaphase, where sister chromatids line up in the center of the cell, anaphase, where sister chromatids separate, and telophase, where the cell with 2 nuclei starts breaking the nucleus into two daughter nuclei<sup>38</sup>.



**Figure 4: Excision Repair Diagram<sup>39</sup>**

Figure 4 shows the checkpoints in the cell cycle provide time for cells to repair damaged DNA and initiate senescence or apoptosis, if necessary.

To ensure that the DNA is faithfully replicated, checkpoints throughout the cell cycle are employed. These checkpoints include the G1/S checkpoint, which is initiated in response to DNA damage and prevents replication of damaged DNA, G2/M, which prevents cells from dividing with damaged DNA, and the spindle assembly checkpoint (SAC), which ensures correct chromosome segregation, occurring between metaphase and anaphase (in case of defects in microtubules or kinetochores). The most significant checkpoint in cancerous cells is the G2/M checkpoint as it provides the cell with time to fully repair any damaged DNA before the cell proceeds to the mitotic phase and thus, stops the proliferation of any harmful damaged DNA. With continuous DNA damage, the cells cannot keep up with DNA repair. UV radiation disrupts bonding in the double-stranded DNA, where side-to-side base pairing occurs, creating a kink in the DNA. In cases like prolonged exposure to DNA, these cyclobutane dimers accumulate, and thus cancerous mutations result due to excision repair not being rapid enough to repair all the DNA damage from UV radiation<sup>40</sup>.

Binding Proteins such as DNA and RNA binding proteins (DRBPs) play a major role in the excision repair process (Figure 4)<sup>41</sup>. Endonucleases are bacterial enzymes that cleave double-stranded DNA and are very specific in their cleavage activity of damaged DNA<sup>42</sup>. DNA Polymerase is then highly significant in the duplication of a cell's DNA, where the polymerase synthesizes nucleotide bases by complementary base pairing to the template strand to repair the damaged DNA<sup>43</sup>. Finally, DNA ligase helps to maintain genomic integrity by joining breaks in the DNA backbone and ligating the new DNA strand into place, thus repairing the damaged DNA<sup>44</sup>.

## G. Role of Melanin Against UV Radiation

Melanin is essential as it plays a crucial role in protecting the body from the harmful effects of UV radiation. Melanin is a pigment primarily found in the skin, hair, and eyes of humans and animals. Melanin is produced and distributed by specialized cells called melanocytes, which synthesize the pigment melanin within melanosomes<sup>45</sup>. Melanocytes are wrapped around a small round organelle called melanosomes whose function is to synthesize and store melanin pigments. When melanosomes are mature, the melanosomes are transported from the melanocytes to keratinocytes, which are the major cell types of the epidermis, and they move to the epidermis layer of the skin. When it is distributed to the keratinocytes, it is retained and provides hair and skin color. Melanocytes are primarily found in the epidermis, the outermost layer of the skin, and are interspersed among the basal layer of skin cells in the keratinocytes. When exposed to UV radiation, melanocytes respond by producing and transferring melanin to the neighboring keratinocytes, providing the skin with its characteristic pigmentation<sup>46</sup>.

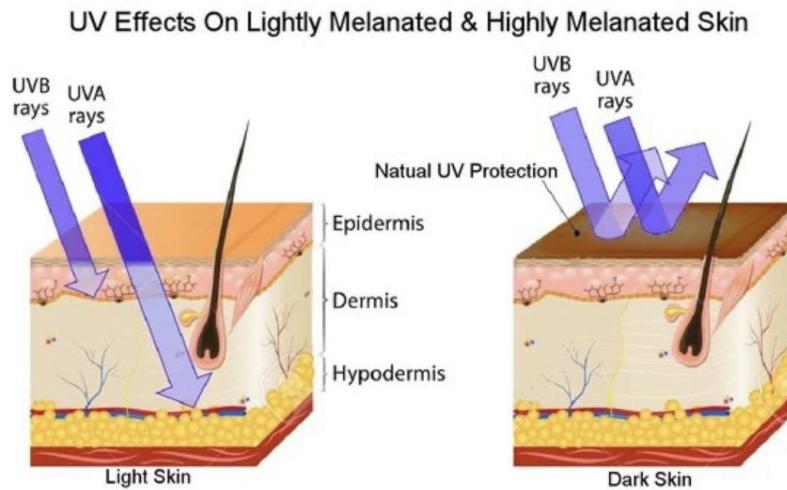
Melanin has a critical role in protecting the body from the harmful effects of UV radiation from the sun. There are three main types of melanin: eumelanin, pheomelanin, and neuromelanin, each with different chemical structures and functions. The most common melanin is eumelanin, which provides the dark colors of the skin, eyes, and hair. Pheomelanin provides pinkish colors such as those found in lips. Eumelanin and pheomelanin are both located in the basal layer of the epidermis. Neuromelanin, however, is different, in that it provides a grayish color to neurons<sup>45</sup>.

The primary role of all forms of melanin is to shield the skin from the damaging effects of UV radiation. When the skin is exposed to sunlight, UV rays can penetrate the outer layer, the epidermis, and reach deeper layers where cells and DNA are located. Therefore, the more UV exposure a person has, these UV rays can cause DNA damage, leading to mutations that may eventually result in skin cancer. Melanin works as a "natural sunscreen" by absorbing and scattering UV radiation, thereby reducing its penetration into the skin. This process helps to minimize DNA damage, reducing the risk of skin cancer development. The more melanin present in the skin, the greater the level of protection it provides against the harmful effects of the sun. Therefore, those with more melanin have a smaller risk of getting skin cancer compared to those with less melanin<sup>47</sup>.

Melanin is also responsible for the color of the skin, hair, and eyes. The type and amount of melanin produced by melanocytes determine an individual's skin tone. This makes both melanin and melanocytes important because the amount and type of melanin produced by melanocytes influence an individual's skin color. People with higher levels of eumelanin typically have darker skin, while those with higher levels of pheomelanin tend to have lighter skin. Melanin is also responsible for the color of the eyes and hair, protecting these sensitive structures from UV-induced damage. In the eyes, melanin is found in the iris and determines eye color. In hair, melanin is synthesized and deposited in the hair shaft, giving it its characteristic color<sup>45</sup>.

While almost everyone has the same amount of melanocytes, the distribution and amount of melanin in the skin can vary significantly among individuals due to genetic factors and environmental influences. One of

the factors that can affect melanin production is sun exposure. UV radiation stimulates melanocytes to produce more melanin, leading to tanning as a protective response against further UV damage (Figure 5). Another factor that can affect melanin production is genetics. Different genetic backgrounds can determine the type and amount of melanin produced, influencing an individual's skin, hair, and eye color<sup>45</sup>. Along with genetics, hormonal fluctuations, such as those occurring during pregnancy, can influence melanin production, leading to skin pigmentation changes known as melasma or "pregnancy mask." Lastly, age is a factor that can affect melanin production because as people age, the production of melanin may decrease, resulting in gray hair and reduced efficiency in protecting the skin against UV radiation<sup>48</sup>.



**Figure 5: UV Effect On Lightly Melanated and Highly Melanated Skin<sup>49</sup>**

Melanin can also act as an antioxidant by neutralizing free radicals, which are highly reactive molecules that can cause cellular damage and contribute to aging and disease. Reactive Oxygen Species (ROS) serves as a cell signaling for biological processes. Oxygen is abundant within cells<sup>50</sup>. It easily accepts free electrons produced by the normal oxidative metabolism in the cell, creating ROS. In addition to this, there is a delicate balance between ROS and the antioxidant defenses that protect the cells from becoming a disease. One of these antioxidant defenses is melanin. However, when exposing melanin to too much UV rays, this will not only harm someone's skin cells, it can also lead to further diseases related to ROS.

It is important to note that while melanin provides some level of natural protection against UV radiation, it's not an absolute defense. Melanin can only absorb 50 - 75% of UV rays<sup>21</sup>. While melanin is crucial for safeguarding the skin, imbalances in melanin production or other genetic factors can lead to skin disorders and diseases. When UV rays exceed the protection from melanin it leads to sunburn, which can mark the beginning of something worse.

## H. SPF-containing Products

Due to the severity of the skin conditions described, there is a range of preventative measures that can be taken. The CDC recommends the Slip, Slop, Slap, and Wrap! Method, which is as follows: "Slip on a shirt, slop on sunscreen, slap on a hat, and wrap on sunglasses"<sup>51</sup>. While there are a wide variety of preventative measures against ultraviolet exposure, the focus of this research project will be on the efficacy of sunscreen as a skin protectant. Several contributing factors influence the effectiveness of sunscreen. Factors such as

the sun protection factor (SPF), the spectrum range, the form of the sunscreen, and the active ingredients affect how well it can protect against the damaging effects of ultraviolet radiation.

SPF is related to the amount of sun exposure and indicates the amount of light that can penetrate the skin's surface. For example, an SPF of 5 indicates that only  $\frac{1}{5}$  of UV rays will be absorbed by the skin, while 100 SPF means that 1/100 of the rays will be absorbed. A higher number for the SPF indicates that fewer photons will reach the skin. SPFs vary, generally ranging from 5 to 100 SPF<sup>52</sup>. The FDA recommends using sunscreen with an SPF value of at least 15<sup>52</sup>, while dermatologists recommend using an SPF of at least 30<sup>53</sup>.

Sunscreens also differ in the type of rays that they protect against. They can be classified as either broad or narrow spectrum. Broad spectrum sunscreen protects the skin from both UVA and UVB rays, as opposed to narrow spectrum which only protects against one type, usually UVB rays. Generally, when a sunscreen does not state what spectrum it protects against, it is a narrow-spectrum sunscreen. As a result, the FDA recommends looking for sunscreens with a broad spectrum, as these provide a greater level of protection<sup>54</sup>.

In addition, it is recommended that sunscreen be applied every 2 hours because its effects diminish over time. Even sunscreens that are labeled as "water resistant" eventually wear off in 40 to 80 minutes<sup>52</sup>. There are also different methods of applying sunscreen. Lotions, sprays, and creams are just some of the many different application methods. Spray sunscreens are easier for accessing hard-to-reach areas, but they may come with a cost due to the risk of inhaling the sunscreen or its propellant. For sprays, to achieve an adequate amount, each body area should be sprayed for at least 6 seconds<sup>55</sup>. Since lotions are thicker in consistency, it is easier to ensure that individual body parts are protected. However, it is ultimately up to the personal preference of each individual to choose which type of sunscreen they want to use.

Furthermore, there are two main classifications for the ingredients in sunscreen: organic and inorganic. Inorganic sunscreens are physical blockers that are designed to reflect the rays off of the skin. They typically contain either zinc oxide or titanium dioxide. Zinc oxide is generally better at reflecting the UVA rays, while titanium oxide is generally better at reflecting the UVB rays. As a result, titanium oxide tends to be more effective, since it can block the rays that are most frequently found to cause skin cancer<sup>56</sup>. These sunscreens are oftentimes bright white because of these ingredients. Recently though, scientists have been using a technique called micronizing to break up these ingredients, resulting in sunscreen that is less white and greasy. These nanoparticles have not been found to penetrate the skin, and there are no known adverse effects of using these sunscreens<sup>57</sup>. Organic sunscreens, on the other hand, use chemical-blocking agents to absorb the sun's rays. They typically contain active ingredients such as oxybenzone, avobenzone, octisalate, octocrylene, and aminobenzoic acid<sup>58</sup>. While there still needs to be further testing, some studies have shown that these ingredients are absorbed into the bloodstream after only a single use<sup>59</sup>. In addition, some studies with animal and human cells have shown that oxybenzone may be a hormone disruptor, impacting the thyroid hormone, testosterone level, and kidney function. However, this tends to be at low levels<sup>60</sup>. Specifically, one study illustrated that when 30,000 parts per million of oxybenzone were added to rat feed, there was an increase in lesions in the kidney and the severity of thyroid gland ultimobranchial cysts<sup>61</sup>. However, current research is not adequate to support a causal relationship. In addition, every active ingredient found in sunscreen is FDA-approved thus far, so there does not seem to be an apparent danger with using chemical sunscreens on humans. However, places such as Hawaii and Key West have banned oxybenzone due to its toxic effects on marine ecosystems<sup>60</sup>. Despite these negative aspects of sunscreen, using any form of sunscreen is still more beneficial than not using any protection at all.

Individuals will want to take into account these factors when choosing their sunscreen, along with their skin phenotype. There are six main classifications of skin phenotype: phenotype I - individuals who always burn and never tan, phenotype II - individuals who experience minimal tanning but easily burn, phenotype III -

individuals who have moderate tanning and burning, phenotype IV - individuals who have minimal burning and moderate tanning, phenotype V - darker skinned individuals who rarely burn but always tan, and phenotype VI - darker skinned individuals who never burn but tan profusely<sup>21</sup>. For example, individuals with phenotype I may have to reapply sunscreen more often than individuals with phenotype VI because of their fairer complexion.

To protect against the damaging rays of the sun, individuals will want to consider all of these factors when choosing which sunscreen and other protective measures to take.

## I. *Saccharomyces cerevisiae*

In this examination of DNA damage upon UV exposure, the yeast *Saccharomyces cerevisiae* was implemented as the model organism. As a single-celled eukaryote belonging to the Kingdom Fungi, it has served as a source to build detailed knowledge regarding phases, structures, and functions of scientific processes including cell division, metabolism, protein folding, apoptosis, and a multitude of signaling pathways<sup>62,63</sup>.

*S. cerevisiae* was the first eukaryotic genome to be completely sequenced in 1996<sup>63</sup>, and the similarities between proteins encoded in both yeast and mammals had been studied long before that. One such example includes the study of the human *Ras* proto-oncogene, for which homologs of *RAS1* and *RAS2* were identified in *S. cerevisiae*. This was one of the earliest signs of yeast and human genome homology, and the extensive studies performed on the yeast *RAS1* and *RAS2* genes have contributed to the understanding of the human *Ras* gene, specifically further comprehension of the *Ras* signaling pathway and cell growth. Since then, the genome of *S. cerevisiae* has been sequenced for homologs and manipulated for further research on several disease-causing genes. For instance, the gene for Werner's syndrome, a genetic and progressive disorder that displays symptoms such as premature aging in mammals, has been found to have a homolog gene, *SGS1*, in yeast that is responsible for reduced life span. This mutant gene in *S. cerevisiae* played a key role in furthering the genetic understanding of Werner's syndrome and its symptoms through convenient genomic analysis because yeast can be studied with rapid and ethical methods<sup>64</sup>. Our experiments using *S. cerevisiae* similarly aim to better understand the DNA damage caused by UV radiation by studying the variable on *S. cerevisiae*.

Extensive gene comparison has found the homology between human and yeast DNA to be between 20% and 30%<sup>65</sup>; this makes the organism a good candidate to undergo gene manipulation that can further knowledge of the human genome. In addition to this substantial sequence homology, yeast is also very easy to grow, and cultures can be grown overnight with a doubling time of approximately 90 minutes<sup>65</sup>. Yeast is simple yet effective, they are easy to grow and removes the ethical aspect of experimentation while being very susceptible to damage from UV radiation. Yeast thrives in YEPD. At an ideal temperature of 32°C, yeast replicates in a manner much more rapidly than mammalian cells, making them an optimal candidate for experiments requiring rapid growth and numerous trials. They are also inexpensive, well-understood, and easily maintained in the lab (Figure 6).

Because this study examines the DNA damage caused by ultraviolet radiation, this repair process must be comparable in yeast and humans such that we can better understand DNA damage in humans. Thus, it is important to note the mechanism by which both *S. cerevisiae* and human cells repair DNA damage is homologous, a process known as nucleotide excision repair (NER). The earliest knowledge of NER as a DNA damage response was identified by studies conducted on *S. cerevisiae*; it was identified as mutations in NER-associated genes abolished UV radiation damage repair in non-cycling cells. This process is understood to extract the damaged oligonucleotide and fill the resulting gap through unscheduled DNA synthesis. NER is now recognized as the main mechanism to remove DNA lesions post-UV exposure and

several environmental mutagens<sup>65</sup>. Because this process is highly conserved between *S. cerevisiae* and humans, yeast is a suitable candidate to investigate this process and how it is affected by variables such as time of exposure, treatments, and prevention methods.



**Figure 6: *Saccharomyces cerevisiae* on an agar plate<sup>66</sup>**

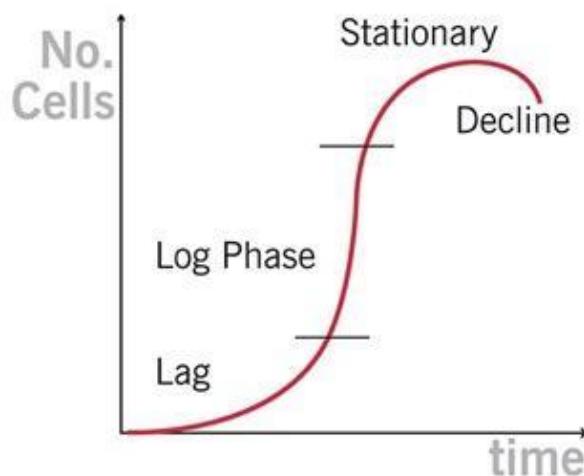
## **II. Methods**

### **A. Yeast Viability Assay**

Yeast viability refers to the ability of how *S. cerevisiae* cultures to grow, reproduce, and interact with their environment. In this case, the extent to which cells were capable of surviving in the UV hood with differing SPF products and times was evaluated.

#### **1. Stable Conditions for *S. cerevisiae***

As the model organism of this study, it is important to provide stable conditions for the growth of yeast. It is stable at a temperature of 32°C when oxygen is present. The strain implemented in this study, BY4741, thrives in a growth media called YEPD: yeast extract peptone dextrose. This medium provides a source of protein and glucose for yeast cells. They then can grow through a process called budding where a small bud grows on the parent cell before separating to become a new cell. *S. cerevisiae* are very susceptible to damage from UV radiation. Yeast go through various growth stages, as shown in Figure 7.



**Figure 7: *Saccharomyces cerevisiae* Growth Curve<sup>67</sup>**

Figure 7 shows the growing stages of yeast as time progresses.

## 2. Serial Dilutions

To successfully analyze the effects of UV radiation on *S. cerevisiae*, 10-fold serial dilutions of *S. cerevisiae* colonies were created from an initial undiluted solution. To create the serial dilutions, 1.0 mL of the undiluted solution is added to 9.0 mL of sterile water and mixed to create a  $10^{-1}$  *S. cerevisiae* solution. After creating the  $10^{-1}$  *S. cerevisiae* solution, 1.0 mL of the  $10^{-1}$  *S. cerevisiae* solution was transferred to 9.0 mL of sterile water and mixed to create a  $10^{-2}$  *S. cerevisiae* solution. This process was repeated until  $10^{-5}$  *S. cerevisiae* dilutions were created. 0.1 mL of cells from both  $10^{-4}$  and  $10^{-5}$  *S. cerevisiae* dilution were spread-plated (see section 2), irradiated, and incubated for 48 hours at 32°C. After the plates were incubated, the resulting *S. cerevisiae* colonies could be directly counted.

## 3. Spread Plate

A cell spreader or plate spreader is a manual tool used in biology laboratories and in any related fields to spread cells and bacteria on a culture plate. To use a cell spreader, it needs to be immersed in ethanol to prevent contamination and maintain cleanliness. After dipping the cell spreader in ethanol, the spreader needs to pass through the flame of a Bunsen burner to complete the sterilization process. Once it is cooled off, it can be used to spread the liquid yeast all around the plate. Furthermore, it is important that after spreading the plate, the tool is resterilized as above. The same process is repeated until every plate is completely spread.

## 4. SPF-containing Product Application

Various SPF-containing products were tested for efficacy on the *S. cerevisiae* cultures as they were irradiated under the UV hood. To apply the specific SPF-containing product, 0.5 g of the desired product was measured and placed centrally on plastic wrap. The plastic wrap was then placed over the *S. cerevisiae* petri dish culture after removing the lid, leaving the side with the SPF-containing product on the outer surface. To ensure that the *S. cerevisiae* was equally protected from UV radiation under the UV hood, the remaining SPF-containing product was evenly distributed across the outer surface of the plastic wrap.

## 5. UV-Radiation & Plate Assessment

To measure the effect of UV Irradiation on the yeast, first, determine a series of time intervals to which the yeast will be exposed. Samples were prepared as in sections 1 and 2 to dilute and spread plate yeast, followed by covering with SPF-containing product, as described in section 3. Both the sunscreen-protected and unprotected yeast cultures were placed under the UV hood for the selected time interval. Once their selected time intervals were complete, the radiated yeast cultures were placed in the incubator at 32°C and left for 48 hours. After the plates were incubated, the number of yeast colonies was counted in each plate to determine the viable count. Viable count is a method used to determine the number of living cells in a cell culture. To count the yeast, it is important to take in mind that there could be contamination in the plate as well. Yeast can be identified by its color: if the colony seems white and creamy, it means that it's yeast, and as for contamination, it has a pinkish color, also yeast is a little inflated on the top, and contamination is flat. If there were too many yeast in a plate (>200 colonies), the plates were divided into quadrants, a single quadrant was counted, and the total was multiplied by the number of quadrants to get an estimate of how many yeast colonies are on the plate. A viability curve of the surviving yeast was created indicating the percent survivors per each condition. Then, for each condition, the viable count was divided by the viable count of the no-UV control plate and multiplied by 100 (see Figure 8). This will give the percentage of the total yeast in the culture that survived exposure for each condition.

$$\text{Percent viability} = \frac{\text{viability count at } t=n}{\text{viability count at } t=0} \times 100$$

**Figure 8: Viability Count Calculation**

After calculating the viability percentages for all time intervals and plates, a percent viability graph was created considering time as the x-axis and percent viability as the y-axis.

## B. Nucleosome Fragmentation Assay

Nucleosome fragmentation assay assesses the viability and signs of apoptosis. Apoptosis can be induced by irreparable DNA mutations, such as those caused by UV radiation. During apoptosis, endogenous endonucleases cut DNA into small fragments. Since the DNA becomes fragmented, proteins cannot be produced correctly in the cell, and cell death is induced. The assay was used to specifically evaluate cells for signs of apoptosis through the detection of randomly sized chromosomal DNA fragments, which is a result of the release of endogenous endonucleases. Using gel electrophoresis, these fragments were identified and assessed for possible yeast DNA damage due to UV exposure. Since UV damage to DNA directly causes apoptosis, there is a direct correlation between the amount of apoptosis and the extent to which UV radiation has caused DNA damage.

### 1. Exposure of Yeast to UV Light and Genomic DNA Extraction

Eight plates were prepared with two conical tubes of the 50 mL yeast, strain BY4741, and were centrifuged at 5,000 rpm for 10 minutes. The supernatant was discarded and the pellets were resuspended in 25 mL of sterile water. 5 mL of yeast was transferred into eight Petri dishes. The eight plates were prepared with a set of four being SPF-positive and the other half being SPF-negative. The SPF-positive group was the experimental group that had SPF-containing product coverage, which was applied using the method mentioned in section 4, while SPF-negative were plates that did not receive SPF application, and only had a plastic wrap when placed under UV radiation. One pair of the Petri dishes were set as the control where  $t = 0$  meaning that there was no UV exposure. The rest of the pairs were placed under UV light for the preselected time points. Once all the plates were taken out of the UV hood, they were placed in an incubator

at 32°C for 30 minutes, allowing the yeast to attempt to repair its DNA damage. Then, the yeast culture was transferred into conical tubes, rinsed with an equal volume of sterile water, and centrifuged at 4000 rpm for 10 minutes. The supernatant was discarded and the pellets were kept in a -80°C freezer.

The genomic DNA was extracted using the Yeast DNA extraction kit (Thermofisher). In brief, cell pellets were defrosted and resuspended in 750 µL of Y-PER (Yeast Protein Extraction Reagent). Yeast was transferred to an Eppendorf tube, and incubated for 10 minutes at 65°C. The tubes were centrifuged for 10 minutes at 4,000 rpm, and the supernatant was discarded. Each pellet was resuspended in 400 µL of DNA-releasing reagent A and reagent B and incubated in the heat block at 65°C for 10 minutes. Then, 200 µL of protein removal reagent was added to each tube, inverted to mix, and centrifuged at max speed for 5 minutes. The supernatant was transferred to a new Eppendorf tube and 600 µL of isopropanol was added and mixed by inversion. The tubes were centrifuged at max speed for 10 minutes, the supernatant was discarded, and 1 mL of ethanol was added. After, the tube was centrifuged at max speed for 1 min, the supernatant was discarded. Then, the tube was left open overnight to completely dry the pellet. The dried pellet was resuspended in 50 µL of TE buffer.

## 2. Spectrophotometry of Yeast Cells

Spectrophotometry was used to determine the absorbance (wavelength), concentration, and purity of the yeast cell samples. The spectrophotometer used in the experiment obtained absorbance of light values in the range of 200-400 nm, and the values at 230 nm, 260 nm, and 280 nm were used for analysis. The value at 230 nm represented the absorbance of peptide bonds in proteins, the value at 260 nm represented nucleic acid absorbance, and 280 nm represented the absorbance of amino acids in the protein. Thus, if the 230 and 280 nm readings on the spectrophotometer were observed to be too high, this would signify that there is too much protein contamination. For the DNA samples to be pure, the calculation performed would be the  $\text{Abs}_{260}/\text{Abs}_{280}$ ; if the ratio was between 1.7-1.9, the sample was pure. The calculation for the concentration of the DNA samples was 50 µg/mL multiplied by the  $\text{Abs}_{260}$ .

## 3. Gel Electrophoresis

To visualize nuclear fragmentation, gel electrophoresis was used to separate genomic DNA by size. Gel electrophoresis applies an electric current to an agarose gel to attract negatively-charged DNA to its positive electrode. When DNA is added to the wells and a current is applied, smaller fragments will be able to travel further through the agarose matrix, while larger fragments will not travel as far. Ethidium bromide (EtBr), when added, allows us to see the fragmentation upon UV irradiation.

Extracted genomic DNA was added to a 0.8% agarose gel containing EtBr. Using the concentration data from spectrophotometry, the volume of DNA solution required for each yeast DNA sample was calculated to maintain a total of 500 ng of DNA. 2 µl of 10X loading dye was added, in addition to the calculated amount of DNA. This calculated amount of DNA is at a maximum of 18 µl. Any additional volume of ddH<sub>2</sub>O was added to ensure that each sample is 20 µl. Reagents were centrifuged three times with manual flicking in between. Mixed samples were loaded into the gel followed by a 1KB DNA ladder. Gels were run at 100V for approximately 2 hours.

## C. Indirect Immunofluorescence

The process of indirect immunofluorescence allows for the staining of a target of interest using a primary antibody to detect and a fluorescent-tagged secondary antibody for identification.

### 1. Preparation of Yeast Cells

5 mL of yeast strain BY4741 and *mcm2 DENQ* culture (a kind gift from Anthony Schwacha, University of Pittsburgh) were grown and diluted by monitoring OD<sub>600</sub> until the yeast cells were in log phase (OD = 0.5 - 1). The yeast cells were then transferred into a petri dish. At this point, two versions of the experiment were carried out. Experiment A focused on the presence of SPF, comparing a +SPF and -SPF condition at a single time point. Experiment B focused on various UV exposure times of the BY4741 yeast cells, all irradiated without SPF protection. For each of these experiments, the *mcm2 DENQ* culture was used as a positive control to illustrate the DNA damage. Other than the t = 0 plates, the Petri dishes containing the cells were placed in a UV hood without lids to induce mutations.

After the yeast was exposed to the UV to their designated time points, 1 mL of the yeast culture was centrifuged for 1 minute at 13,000 rpm and the supernatant was removed. Another 1 mL of irradiated yeast cells was added to the same Eppendorf tube, centrifuged, and the supernatant was removed. Yeast pellets were resuspended in 1 mL of potassium phosphate buffer (KPO<sub>4</sub>) (2.74 mL 1M KH<sub>2</sub>PO<sub>4</sub>, 1.26 mL 1M K<sub>2</sub>HPO<sub>4</sub>, 1.5% BSA, 0.5% Tween20, 1M MgCl<sub>2</sub> + 0.1% Triton X-100 in 50 mL PBS) containing 4% paraformaldehyde. The tubes were incubated on a platform shaker for 1 hour at room temperature.

Slides were prepared for immunofluorescence by applying 10 uL of 0.1% poly L-lysine solution to each well of the 10-well slide. The poly-L-lysine solution was incubated at room temperature for 10 minutes. Then each well was washed 5 times with 20 uL of sterile water. To ensure accurate results, two wells on the outside of the slide were excluded, so only 6 of the wells were used. This is because at the end of the experiment when we seal the slide with a glass coverslip with nail polish, we want to avoid contaminating the yeast cultures with the polish.

The fixed cells were centrifuged for 1 minute and the supernatant was discarded. Cells were resuspended in 500 uL of KPO<sub>4</sub> buffer, centrifuged and the supernatant was removed. Next, cells were washed in the KPO<sub>4</sub> buffer. The cells were resuspended in 0.1 mL sorbitol buffer (2.2 g sorbitol, 0.271 mL KH<sub>2</sub>PO<sub>4</sub>, 0.126 mL K<sub>2</sub>HPO<sub>4</sub>, and 5 uL MgCl<sub>2</sub> in 10 mL water, pH = 6.5). To remove the cell wall, 5 uL of the glusulase was added to each tube. These tubes were incubated in a heat block for 15 minutes at 30°C and then centrifuged for 1 minute. Cell pellets were washed 3 times in the sorbitol buffer. 20 uL of each of the cells was applied into one of the six wells treated with poly-L-lysine and incubated at room temperature for 10 minutes. The slide was transferred to a methanol chamber for 6 minutes, and an acetone chamber for 30 seconds immediately after. After drying, 20 uL of blocking solution (1.5% BSA, 0.5% Tween20, 0.1% Triton X-100 in PBS) was placed in each well before incubating overnight in a humidified chamber to prevent the cells from drying overnight.

### 2. Incubation with Antibodies

10 uL of a 1:1000 dilution of rabbit-anti-γ-H2AX (Abcam) in PBS was added to each well. This antibody can recognize the phosphorylation mark that shows the DNA is damaged. The solution was then incubated at room temperature overnight in a humid chamber. The next day, the cells were washed with 20 uL of PBS a total of 5 times to remove any excess of the primary antibody. Subsequently, 20 uL of PBS was added and the cells were incubated for 10 minutes at room temperature. For the secondary antibody, 20 uL of

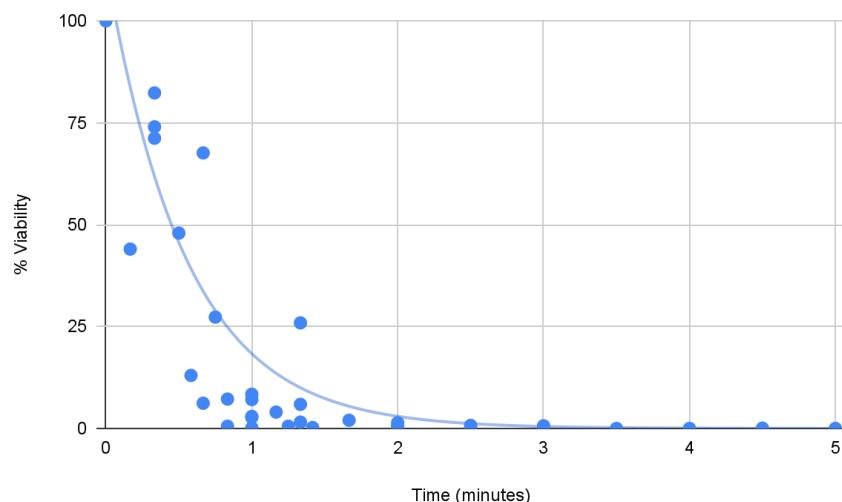
1:500 dilution of goat-anti-rabbit-Cy3 (Abcam) in PBS was added and incubated in the dark at room temperature for one hour. This antibody will bind to the primary antibody's constant region, and the Cy3 marker can illustrate the DNA damage. After the incubation, the cells were washed 5 times with PBS to remove any excess of the secondary antibody. 20  $\mu$ L of PBS was then added to the cells and left to incubate for 10 minutes at room temperature. The excess was pipetted off. Next, 20  $\mu$ L of the DAPI (4', 6-diamidino-2-phenylindole)/fluoromountG dye was added to the wells and incubated for 10 minutes. DAPI is a fluorescent blue stain that stains the nucleus of the cells. After the DAPI was applied, a coverslip was added to the slide. Slight pressure was applied to ensure that air bubbles were removed. The edges of the coverslip were then sealed using clear nail polish and the slides were kept at -20°C until imaging was ready to take place.

### 3. Imaging

The yeast cells were observed under an Epifluorescent microscope and imaged under white light, the DAPI (blue), and the Cy3 (red) filters. Each image was put under the auto-expose setting to see the fluorescent nature of the DAPI and Cy3 in the absence of light. Then, an image for each light setting for each well was downloaded, yielding a total of three images for each of the six wells. Finally, a program called ImageJ was used to false-color the DAPI pictures as blue and the Cy3 as red. Two images were then overlaid on top of each other to see where there was both viable DNA (shown in the presence of blue/DAPI) and DNA damage (presence of red).

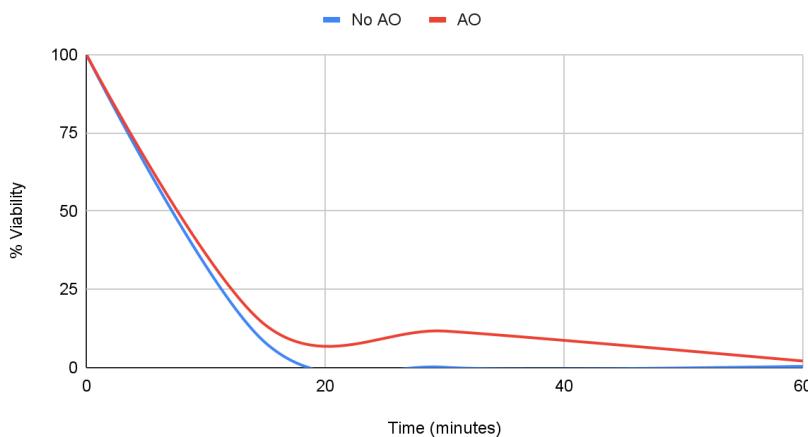
## III. Results

### A. Viability Assays



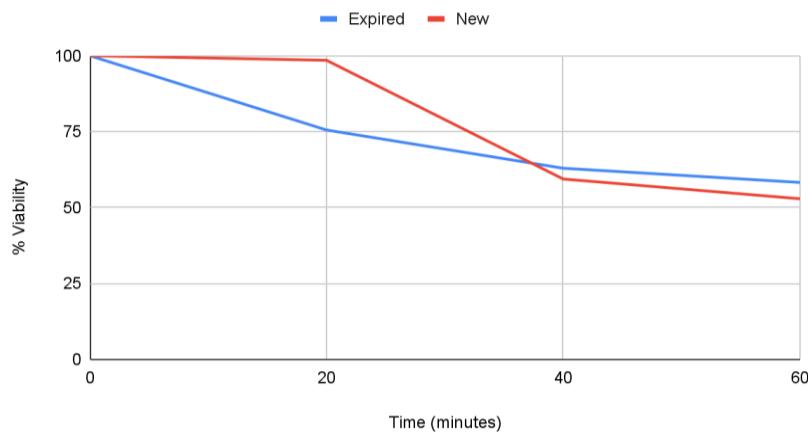
**Figure 9: Killing Curve of Wild Type Yeast with No SPF**

Figure 9 represents the percent viability for yeast cells that were exposed to UV light without any SPF protection.



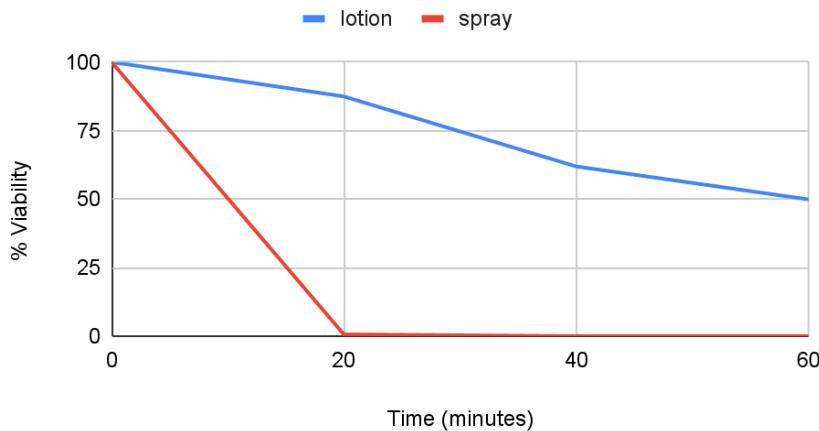
**Figure 10: Killing Curve of Antioxidant vs. No-Antioxidant Sunscreen**

Figure 10 represents the comparison of sunscreen containing antioxidants as an ingredient (red) vs. sunscreen that does not have antioxidants (blue) when irradiated from time points 0 to 60 minutes.



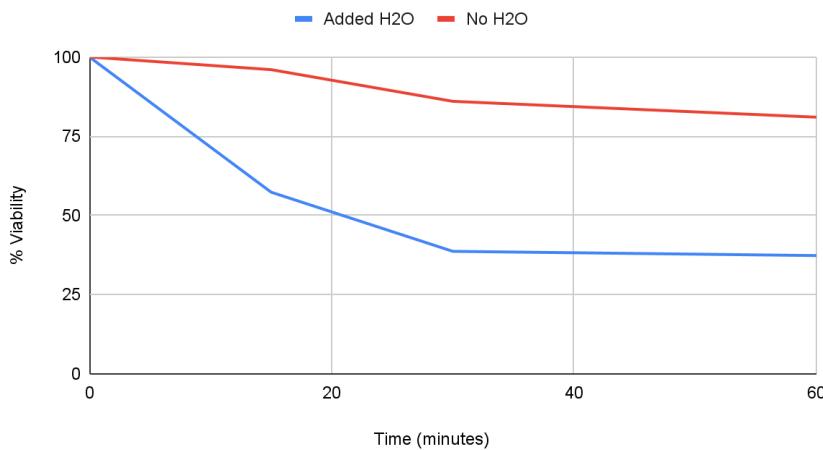
**Figure 11: Killing Curve of Expired vs. Non-Expired Lotion Sunscreens**

Figure 11 compares two Coppertone 50 SPF sport sunscreens, one that is expired and one that is not. The new bottle (red line) is set to expire in 2024. The expired bottle (blue line) expired in 2021. Yeast samples were irradiated for 0 to 60 minutes.



**Figure 12: Killing Curve of SPF 50 Lotion vs. Spray**

Figure 12 shows two Coppertone sport sunscreens with an SPF of 50, one being spray (red line) and the other being lotion (blue line). Yeast samples were irradiated for 0 to 60 minutes.



**Figure 13: Killing Curve of Water Resistance - Added H<sub>2</sub>O vs. No H<sub>2</sub>O**

Figure 13 shows the comparison of water resistance sunscreen when diluted with water (blue) versus when the sunscreen was not in contact with water (red). Yeast samples were irradiated for 0 to 60 minutes.

## B. Nucleosome Fragmentation Assay

### 1. Spectrophotometry Calculations of gDNA extractions

Note: +SPF indicates sunscreen is applied and -SPF indicates sunscreen is not applied

**Table 1: Coppertone SPF 50 Sport Lotion**

Experiment 1	Time (min)	Purity	Concentration ( $\mu\text{g/mL}$ )
+SPF	0	1.2357	134.75
+SPF	20	1.4466	77.25
+SPF	40	2.1361	120.05
+SPF	60	1.5011	135.25
-SPF	0	1.1079	149.90
-SPF	20	1.4258	149.35
-SPF	40	2.1768	73.2
-SPF	60	0.9560	62.00

**Table 2: Coppertone SPF 15 Sport Lotion**

Experiment 2	Time (min)	Purity	Concentration ( $\mu\text{g/mL}$ )
+SPF	0	2.2451	80.60
+SPF	10	2.5334	90.95
+SPF	20	1.3969	79.90
+SPF	30	3.2500	34.45
-SPF	0	1.9655	2.85
-SPF	10	1.49020	7.6
-SPF	20	0.86890	25.85
-SPF	30	1.03571	10.25

**Table 3: Coppertone SPF 50 Sport Spray**

Experiment 3	Time (min)	Purity	Concentration ( $\mu\text{g/mL}$ )
-SPF	0	2.0226	31.35
-SPF	10	2.0088	11.35
-SPF	20	11.5000	2.30
-SPF	30	2.0237	25.60
+SPF	0	2.4141	15.45
+SPF	10	1.4546	105.75
+SPF	20	2.0148	20.45
+SPF	30	1.8655	139.35

**Table 4: Coppertone SPF 100 Water Resistant Lotion**

Experiment 4	Time (min)	Purity	Concentration ( $\mu\text{g/mL}$ )
+SPF	0	1.89	67.1
+SPF	50	0.703	49.95
+SPF	55	1.7	45.3
+SPF	60	1.3	40.45
-SPF	0	2.13	102.75
-SPF	50	1.75	124.35
-SPF	55	0.957	48.55
-SPF	60	2.0	119.55

**Table 5: Neutrogena SPF 20 Foundation**

Experiment 5	Time (min)	Purity	Concentration ( $\mu\text{g/mL}$ )
+SPF	0	1.739	126.8
+SPF	10	2.316	64.15
+SPF	20	1.012	49.92
+SPF	30	1.729	153.25
-SPF	0	1.876	135.45
-SPF	10	2.109	126.25
-SPF	20	1.65	144.55
-SPF	30	1.163	155.7

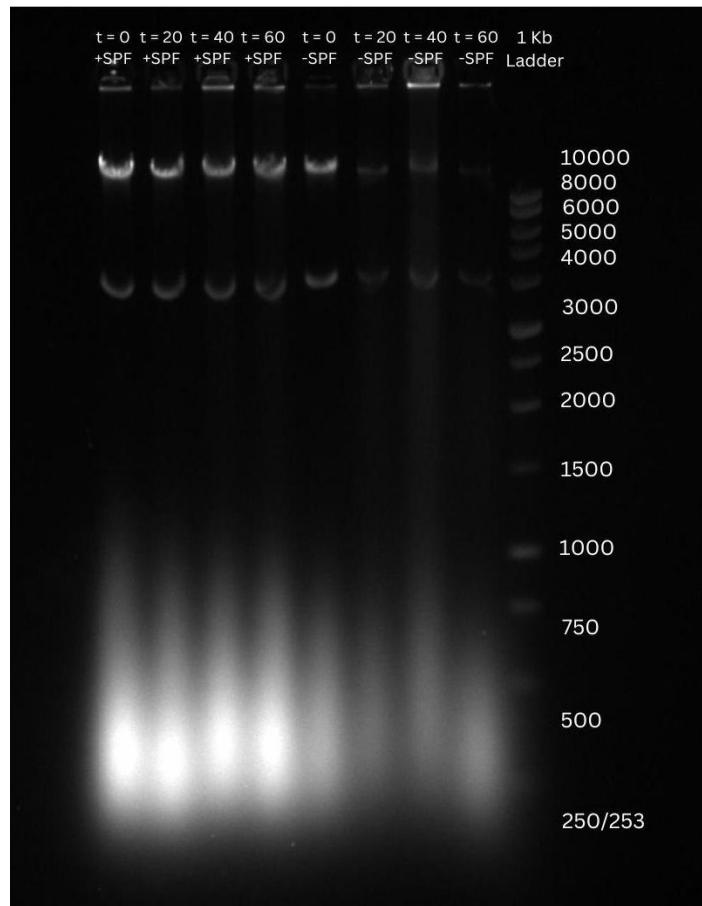
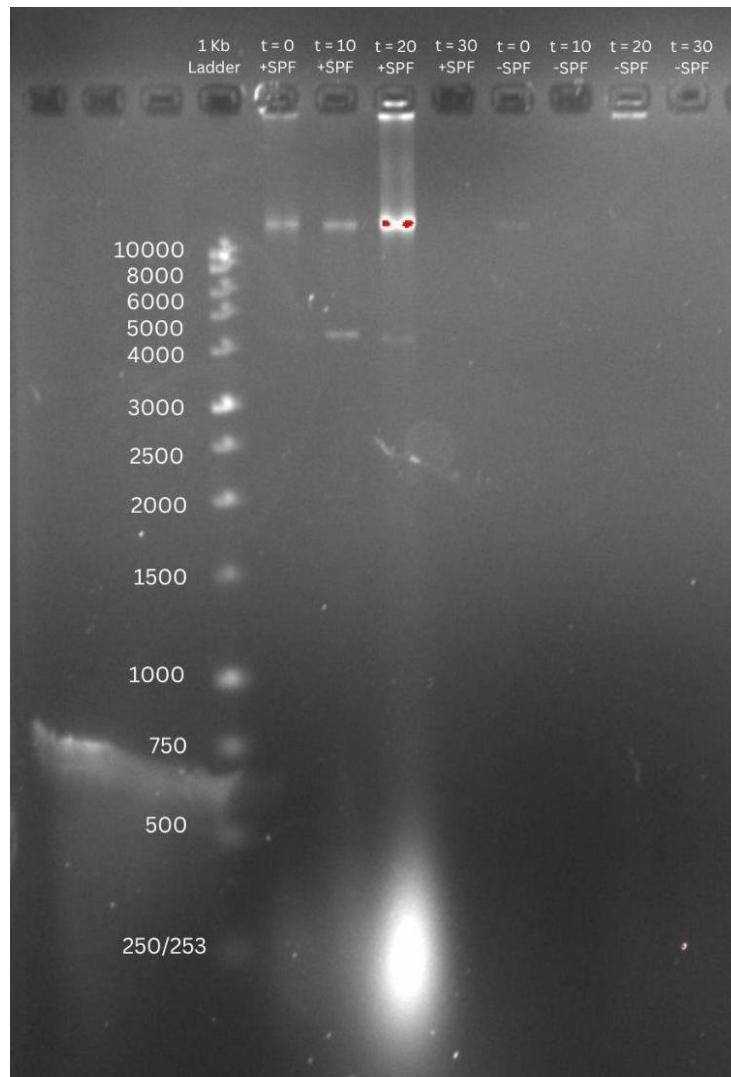
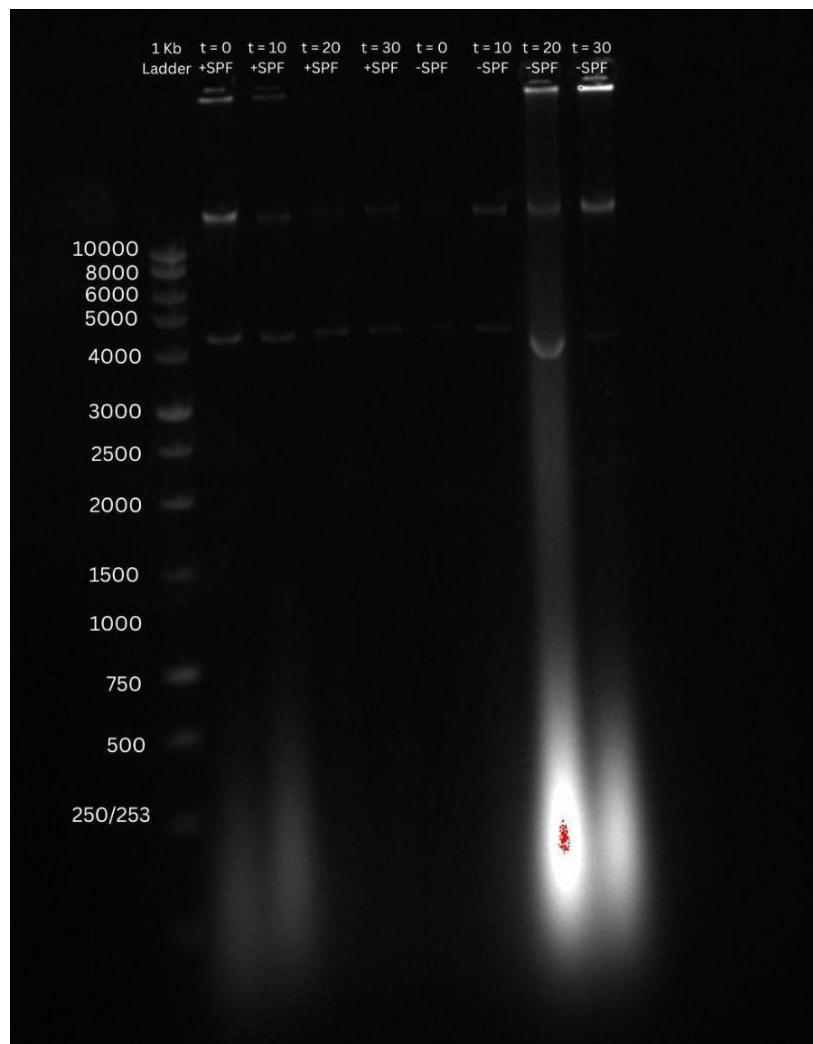
**2. Agarose Gel Electrophoresis Images****Figure 14: Coppertone SPF 50 Sport Lotion (Experiment 1)**

Figure 14 corresponds to Table 1, and it shows the gDNA fragmentation after UV irradiation. A 1 Kb ladder, visible in well 10, is used to distinguish DNA sizes. 8 lanes of gDNA are added lanes 1 - 8. Lanes 1 - 4 contain gDNA extracted from +SPF samples covered with Coppertone SPF 50 sport lotion sunscreen. Lanes 5 - 8 contain gDNA extracted from -SPF samples not covered with sunscreen. Lanes 1 and 5 contain samples t = 0 minutes of UV exposure. Lanes 2 and 6 contain samples t = 20 minutes of UV exposure. Lanes 3 and 7 contain samples t = 40 minutes of UV exposure. Lanes 4 and 8 contain samples t = 60 minutes of UV exposure.



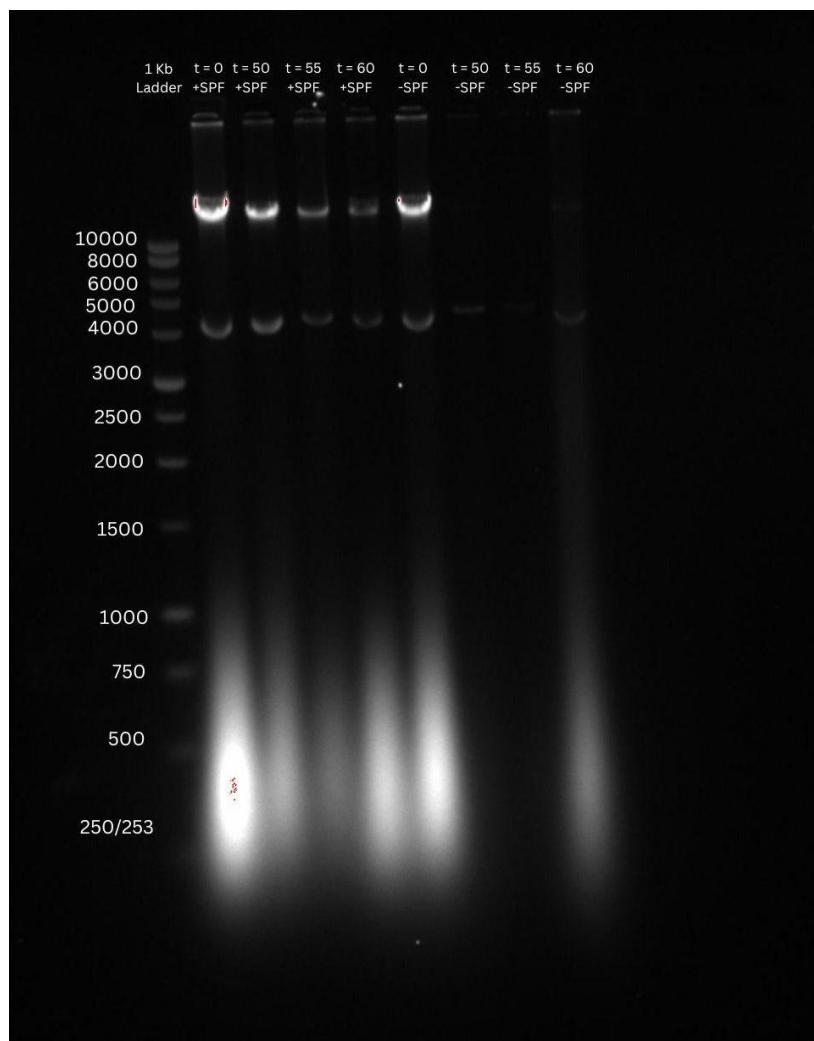
**Figure 15: Coppertone SPF 15 Sport Lotion (Experiment 2)**

Figure 15 corresponds to Table 2, and it shows the gDNA fragmentation after UV irradiation. A 1 Kb ladder, lane 10, is used to distinguish DNA sizes. 8 lanes of gDNA are added to lanes 1 - 8. Lanes 1 - 4 contain gDNA extracted from +SPF samples covered with Coppertone SPF 15 sport lotion sunscreen. Lanes 5 - 8 contain gDNA extracted from -SPF samples not covered with sunscreen. Lanes 1 and 5 contain samples t = 0 minutes of UV exposure. Lanes 2 and 6 contain samples t = 10 minutes of UV exposure. Lanes 3 and 7 contain samples t = 20 minutes of UV exposure. Lanes 4 and 8 contain samples t = 30 minutes of UV exposure.



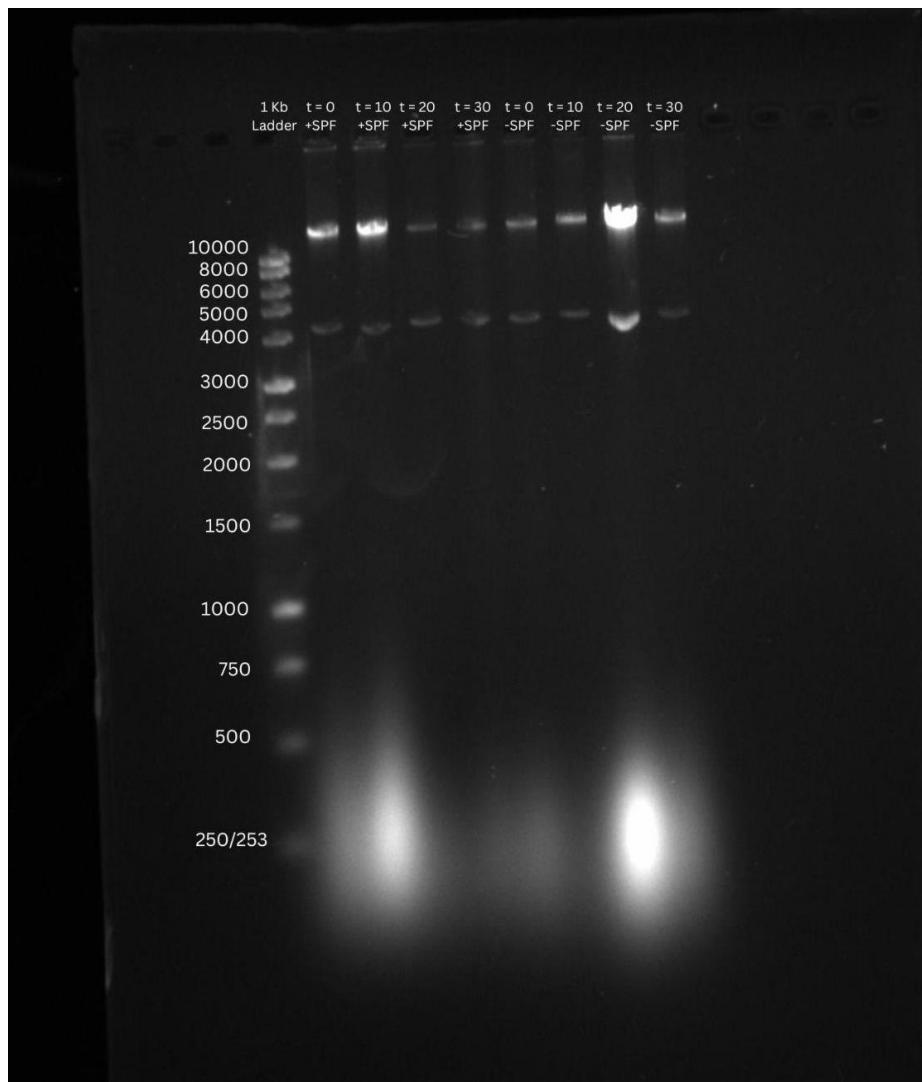
**Figure 16: Coppertone SPF 50 Spray (Experiment 3)**

Figure 16 corresponds to Table 3, and it shows the gDNA fragmentation after UV irradiation. A 1 Kb ladder, visible in the far right well, is used to distinguish DNA sizes. 8 lanes of gDNA are added to lanes 1 - 8. Lanes 1 - 4 contain gDNA extracted from +SPF samples covered with Coppertone SPF 50 Spray. Lanes 5 - 8 contain gDNA extracted from -SPF samples not covered with sunscreen. Lanes 1 and 5 contain samples t = 0 minutes of UV exposure. Lanes 2 and 6 contain samples t = 10 minutes of UV exposure. Lanes 3 and 7 contain samples t = 20 minutes of UV exposure. Lanes 4 and 8 contain samples t = 30 minutes of UV exposure.



**Figure 17: Coppertone SPF 100 Water Resistant Lotion (Experiment 4)**

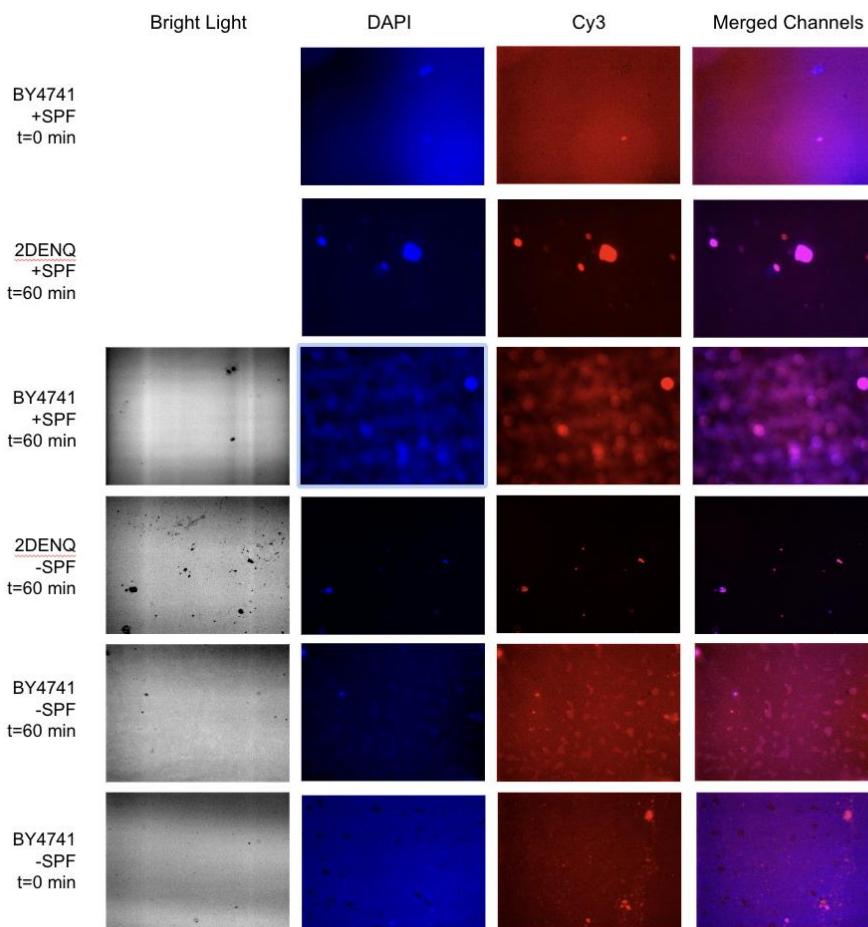
Figure 17 corresponds to Table 4, and it shows the gDNA fragmentation after UV irradiation. A 1 Kb ladder, visible in the far right well, is used to distinguish DNA sizes. 8 lanes of gDNA are added to lanes 1 - 8. Lanes 1 - 4 contain gDNA extracted from +SPF samples covered with SPF 100 sunscreen. Lanes 5 - 8 contain gDNA extracted from -SPF samples not covered with sunscreen. Lanes 1 and 5 contain samples t = 0 minutes of UV exposure. Lanes 2 and 6 contain samples t = 50 minutes of UV exposure. Lanes 3 and 7 contain samples t = 55 minutes of UV exposure. Lanes 4 and 8 contain samples t = 60 minutes of UV exposure.



**Figure 18: Neutrogena SPF 20 Antioxidant Liquid Makeup (Experiment 5)**

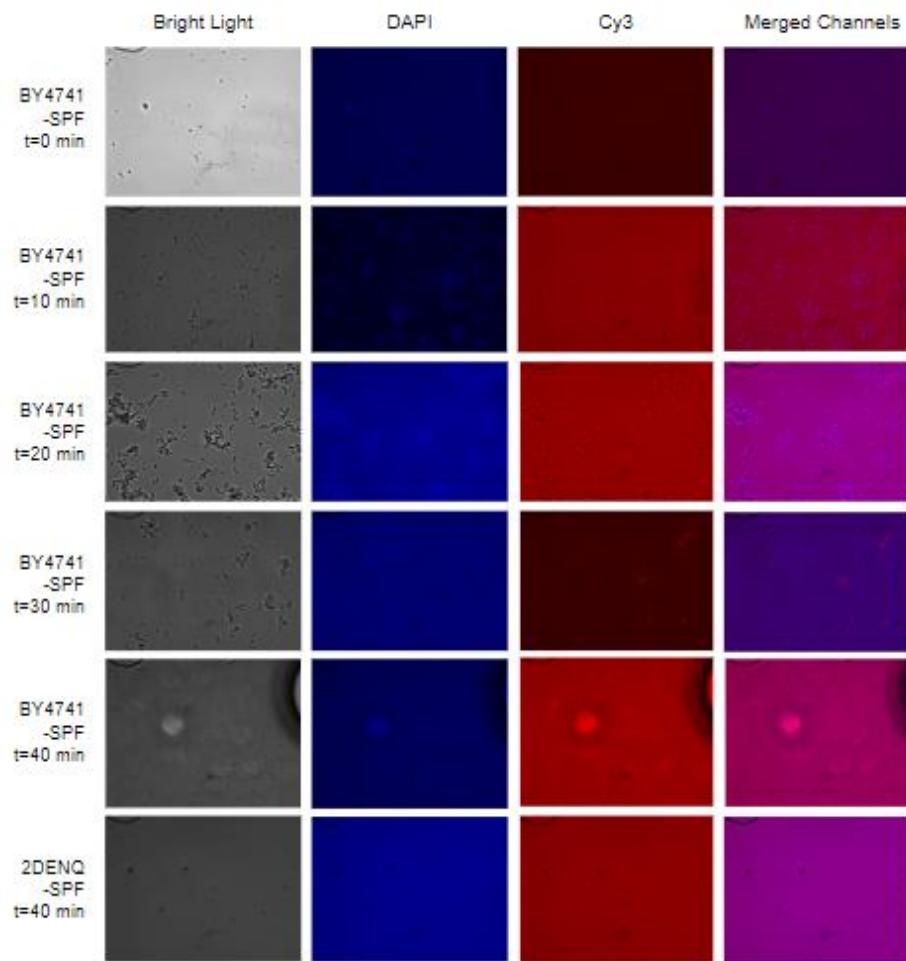
Figure 18 corresponds to Table 5, and it shows the gDNA fragmentation after UV irradiation. A 1 Kb ladder, visible in the far right well, is used to distinguish DNA sizes. 8 lanes of gDNA are added to lanes 1 - 8. Lanes 1 - 4 contain gDNA extracted from +SPF samples covered with Neutrogena SPF 20 antioxidant liquid makeup. Lanes 5 - 8 contain gDNA extracted from -SPF samples not covered with makeup. Lanes 1 and 5 contain samples  $t = 0$  minutes of UV exposure. Lanes 2 and 6 contain samples  $t = 10$  minutes of UV exposure. Lanes 3 and 7 contain samples  $t = 20$  minutes of UV exposure. Lanes 4 and 8 contain samples  $t = 30$  minutes of UV exposure.

### C. Indirect Immunofluorescence Images



**Figure 19: Immunofluorescence Imaging of Eucerin SPF 30 Moisturizer**

Figure 19 shows Immunofluorescence Experiment A which utilized immunofluorescence imaging of yeast cells with and without Eucerin SPF 30 moisturizer after exposure to UV light for the given time intervals. The DAPI staining shows the location of the nuclei, the Cy3 staining shows where cell damage occurred, and the overlap shows where DNA damage occurred inside the cell. Note: it was difficult to visualize the yeast cells with the white light for the BY4741, t = 0, +SPF, and for the 2DENQ, t = 60, +SPF. As a result, white light images were unable to be obtained for those timepoints.



**Figure 20: Immunofluorescence Imaging of Yeast without SPF product**

Figure 20 shows Immunofluorescence Experiment B which used immunofluorescence imaging of yeast cells that were exposed to UV light at different time intervals. The DAPI staining shows the location of the nuclei, the Cy3 staining shows where cell damage occurred, and the overlap shows where DNA damage occurred inside the cell.

## IV. Discussion

### A. Viability Assays

To determine the effect of SPF-containing products on yeast survival, we constructed a killing curve by plotting percent viability against time exposed to UV radiation. Upon being taken out of the incubator after 48 hours, yeast colonies on the agar plate were counted and percent viability was determined by dividing the UV irritated count by the control count ( $t = 0$ ). Figure 9 shows that yeast viability decreased to 0 percent after 180 seconds, which indicates that the total cell death occurred at that time. The total yeast viability assay serves as a control of neutral conditions, proving UV-induced cell death, which we based the rest of our killing curves upon.

Figure 10 represents the comparison of sunscreen containing antioxidants as an ingredient vs. sunscreen that does not have antioxidants in yeast viability after 60 minutes of UV irradiation. The figure shows that yeast protected by antioxidant sunscreen proved to have a slightly higher percent viability throughout the entire 60-minute duration than the yeast protected by sunscreen with no antioxidants. This was especially noticeable during the 20-40 minute mark, where the yeast protected by antioxidants demonstrated 15 percent higher viability than the yeast protected by no antioxidant sunscreen. However, after the 40-minute mark, the two lines seem to even out. This indicates that sunscreen containing antioxidants is only effective within small time intervals.

Figure 11 shows two Coppertone Sport SPF 50 lotions with varying expiration dates. One bottle expired in 2021 and the other has an expiration date of 2024. The non-expired sunscreen did well in protecting the yeast up until the 20-minute mark, where it had a 98% yeast viability compared to a 75% viability of yeast protected by expired sunscreen. After 20 minutes, both sunscreens had similar percent viability. Therefore, we can conclude that expired sunscreen protects yeast similarly to non-expired sunscreen, especially in later timepoints.

Figure 12 shows two Coppertone Sport sunscreens with an SPF of 50, one liquid spray-based and another lotion-based. The yeast protected by spray-on sunscreen had a percent viability of around 0% by the 20-minute mark, while the yeast protected by lotion sunscreen had a percent viability of about 90% at that same time. The lotion continued to protect the yeast throughout the duration of the experiment while the spray continued to have 0% viability of yeast. Therefore, we can conclude that lotion sunscreen offers better protection than spray sunscreen.

Figure 13 shows the comparison of water resistance sunscreen when diluted with water versus when the sunscreen was not in contact with water. As shown, the water resistance sunscreen was not effective when sprayed with water as the percent viability decreased and was significantly lower than when the sunscreen that was not sprayed with water. The percent viability of the yeast protected by sunscreen sprayed with water decreased to 33% at the 30-minute mark, while the percent viability of the yeast protected by sunscreen not sprayed with water stayed steady at 80% starting from 30 minutes. Therefore, the water-resistance sunscreen did not exhibit its claim of effectiveness in protection against UV rays when in contact with water.

### B. Nucleosome Fragmentation Assay

The first step of analyzing the gel electrophoresis was to compare the  $t = 0$  lanes for both the wells with and without SPF-containing products. If the bands looked very similar, it was able to be confirmed that the same amount of DNA was added to the gels, as performed in our calculations from tables 1 to 5. This

enables us to analyze the results of the gels. In Figure 14, we tested the Coppertone 50 SPF Sport Lotion. For the wells with the sunscreen, there are highly visible bands towards the top of the gel. This means that there is a larger DNA fragment in this gel, indicating that this sample had less shearing of genomic DNA. For the wells without the sunscreen, it is difficult to visualize the DNA, meaning that the DNA is broken up into smaller fragments that are spread throughout the lane. This illustrates that there was a significant amount of DNA damage in the lanes where sunscreen was not applied. The bright smear at the bottom of the gel shows RNA, which does not relate to the genomic DNA we attempted to analyze. As a result, this shows that the Coppertone 50 SPF lotion effectively prevented DNA damage. Furthermore, in Figure 17, we can see similar results. The  $t = 0$  lanes are very similar in brightness. In addition, the lanes with the sunscreen all have a distinct band, meaning that there is less DNA damage, while the lanes without clear bands have significant DNA damage. The remaining gels could not be analyzed so therefore the results are inconclusive because there is not a baseline to reference. This will be discussed further in the error analysis section.

### C. Indirect Immunofluorescence

We used indirect immunofluorescence to examine DNA damage to the yeast cells. This was done for both Experiment A and B, allowing us to visualize the damage on DNA induced by UV radiation over differing time points, as well as the presence or absence of sunscreen. In the second column of images, DAPI stained the nuclei of all yeast cells present blue. The next column illustrated the red Cy3 strain, which recognized any DNA damage the cell had undergone. This occurred through the primary antibody rabbit-anti- $\gamma$ -H2AX, which was able to recognize phosphorylated DNA, indicating damage. The last column merges these channels, so that any combined DAPI/Cy3 staining appears as purple, and indicates a cell that was damaged due to radiation. In Experiment A (Figure 19), the DNA damage for +/- SPF was observed. In Experiment B (Figure 20), the DNA damage at several different time points was examined. Both experiments implemented a yeast strain containing a mutation in *mcm2DENQ*, in which we expected to see damaged DNA. This served as a positive control for the rest of our images.

As seen in Figure 19, the yeast cells that were not exposed to any UV radiation had no purple fluorescence, which proves that the DNA damage in other samples must have been solely from the UV radiation. The amount of purple fluorescence in the merged channel was lower for the yeast cells with protection from SPF products. This means that the SPF products were effective in protecting the yeast cells from DNA damage. In the *mcm2DENQ* immunofluorescence, there was significant purple fluorescence, which means that the *mcm2DENQ* sample had significant DNA damage. This supports that the positive control already contained DNA damage and supports that our method could identify DNA damage.

The yeast cells in Figure 20 were not exposed to any UV radiation and had no purple fluorescence, which proves that the DNA damage in other samples must have been solely from the UV radiation. The amount of purple fluorescence in the merged channel increases as time exposed to UV radiation increases. For instance, there is more purple fluorescence in  $t = 20$  min than in  $t = 10$  min. This establishes a direct correlation between the amount of UV radiation and the frequency of DNA damage in the yeast cells. As the yeast cells were exposed to UV radiation for a longer period, more DNA damage accumulated. In the *mcm2DENQ* immunofluorescence, there was significant purple fluorescence, which means that the *mcm2DENQ* sample had significant DNA damage. This supports that the positive control already contained DNA damage.

## V. Error Analysis

Though there were successes in the experiment, it is also important to take into account that there could be some errors that could have affected the results. One of the most important factors that can affect the results of the experiment is contamination. Accidentally reusing pipette tips could have led to contamination, since it has been used in other parts of the experiment. For some viability assays, contamination invalidated results due to difficulty in counting colonies. Another thing to consider is the use of the spreader and not placing it in ethanol between uses. Using ethanol leads to further purification of the spreader, so if ethanol isn't used on the spreader, it can lead to further contamination of the plate.

The immunofluorescence imaging in our study had flaws in several of the group's experiments. This left the images to be inconclusive and too few yeast cells were preserved to derive results from. This may have been sourced by contamination of the wells when left to sit on the workbench at room temperature for 10 minutes or when sealing the slide with nail polish. Errors in addition to poly-L lysine solution may have prohibited yeast cells from sticking to the wells during the washing processes. Additionally, some groups used only four out of the six wells, and another group had broken their slide when imaging, reducing the number of samples able to be analyzed.

Some specific errors in this experiment were due to human error. Some examples of this are improper application of the SPF products, improper sterilization, and miscalculations when counting colonies. One factor that can play a role in contamination is if the cultures of yeast cells come in direct contact with the sunscreen. Some team members had placed the plastic wrap covered with sunscreen on the inside of the plate rather than the outside. If the sunscreen had touched the yeast cultures, it could have led to contamination on the plate, inhibiting yeast growth. Furthermore, when working on the Nuclear Fragmentation Assay, we worked with the DNA of the yeast, so it was vital to change gloves periodically and keep a sterile bench at all times. If proper sterilization was not upheld, DNA could have been degraded, leading to additional shearing and a more visible smear on the gel. Equally important, miscalculations of yeast cultures could occur due to counting contaminated colonies or by counting previously counted plates.

## VI. Future Research

Regarding the future directions of our research, there are several other variables that we would be interested in testing. Because our research outlines problems caused by UV radiation, our next step in research would also be toward preventative methods and solutions. This section explores additional variables that are necessary in furthering our understanding of DNA damage and UV radiation, as well as steps toward solutions.

### A. Long-Term Exposure

During the experimentation process, yeast was exposed to radiation for no more than 1 hour at a time. All the results derived for percent viability and DNA damage only considered the effect of consistent radiation during these shorter periods. Given more time, it may be pertinent to expose yeast colonies to radiation over longer periods in shorter increments before obtaining results. This could mimic the effect of cells' repair mechanisms over time and model the effects of continued UV radiation on DNA. For this form of the experiment, the yeast would have to be preserved over longer periods, and the processes of incubation and radiation may change. This could also highlight the importance of consistently applying sunscreen over long periods.

## B. Protective Clothing

Rather than focusing on the DNA repair mechanisms induced by UV radiation, we can shift our focus to prevention methods in addition to sunscreen. This includes complete protection of skin from UV rays through clothing. Though there are different types of clothing, specifically making clothing that can absorb UV radiation will benefit many people at risk for sunburn. Testing different types of cloth can lead to a better understanding of which clothing can protect the skin. Future researchers can analyze the color, material, Ultraviolet Protection Factor (UPF), and size when looking for clothes to shield our skin from UV rays. Different types of colors, such as dark or light, can absorb different amounts of UV rays. Analyzing the size and fits of clothing can lead to a further understanding of how it can affect the skin with UV radiation. Doing further research about types of clothing and how they can protect the skin from UV rays can lead to a greater understanding of how to defend our skin against it<sup>68</sup>. It can also lead to the development of better pieces of clothing specifically geared to protect certain skin types.

## C. Chemical Ingredients

The sunscreens used in our experiment contain controversial ingredients like oxybenzone, avobenzone, and octisalate. It remains unknown if these ingredients might have adverse effects or any effect to the viability of our yeast cultures. Oxybenzone, in previous studies, has been shown to disrupt hormones in humans due to its rapid absorption rate in the skin. Avobenzone is an organic UV filter that has to be paired with other ingredients to prevent it from breaking down. However, studies have shown that avobenzone can disrupt the endocrine system and cause allergic reactions when it breaks down. Octisalate is another organic UV filter that absorbs through the skin at a rapid pace to the point where the FDA has raised concerns about the high levels found in blood. Further research is necessary in order to determine if the hypothesized toxicity of these chemical ingredients outweighs the long established use of these ingredients. By investigating whether sunscreens without these ingredients are as effective in protecting yeast cells against UV radiation, we may be able to consider the eradication of these ingredients in SPF containing products<sup>69</sup>.

## D. Mutations and Repair Mechanisms

The damage to DNA sequences induced by UV radiation is known and has been demonstrated by our experiments. However, the investigation of these specific mutations in the human genome could be relevant in understanding the mechanisms that lead to detrimental effects of UV radiation. By doing so, we can investigate prevention methods within the specific repair mechanisms. Understanding how these mutations and repair mechanisms are affected by age, environment, long-term exposure, and genetic background may help us develop prevention methods.

## E. Reflected UV Light

Future tests could be conducted on different surface textures and areas to see how the reflection of UV light on a surface affects the radiation we experience. These surfaces include water, snow, ice, and grass. When looking at the percent reflection of UV on snow, 85% of UV rays were reflected. This large percentage indicates that reflective UV light can cause damage relative to the rates of direct UV radiation. However, if we look at the reflection of UV light on various grass surfaces it only reflects around 2.5% of its rays, according to the demographics of the Eccles Health Sciences Library<sup>70</sup>. Future research investigate the different surfaces that reflect light and determine if there is reason for worry regarding the effects of reflected UV radiation. Additionally, this information may be applied to prevent exposure through the manipulation of surfaces and further understanding of how these surfaces regulate detrimental exposure to UV radiation.

## F. Sunscreen Development

Using sunscreens with high levels of SPF is useful for lightly melanated skin, however it may be more essential to reapply every two hours, as outlined in the background of our study. For this reason, in developing sunscreen, increased SPF levels may not be of main consideration. Studies have shown that SPF levels above 100 are not heavily recommended because as SPF levels increase, the amount of coverage increases by less. Double coverage is the most beneficial because it decreases the number of spots that may not be fully covered by the sunscreen. For example, people who applied sunscreen once by a median of 20% missed spots, while a reapplication led to 9% missed spots<sup>71</sup>. Thus, reapplication is a key factor in maximizing the efficacy of sunscreen, and it should be a future goal to spread that awareness. In addition, more research is needed to manufacture water-resistant sunscreen that can fully withstand different water conditions. Some future solutions could have to do with developing more sunscreen that can attach to wet skin, such as those with synthesized fatty acids. Compounds such as cetyl ethylhexanoate are primarily used in hair and skin products, but they could be incorporated into sunscreen to ensure that they are truly effective when under the influence of water. This group of compounds is also reasonably safe when tested on animals, so they can serve as an alternative to the ingredients used currently<sup>72</sup>. Lastly, we could research different components of sunscreen that are both effective at protecting against UV rays and harmless to the environment. Currently, compounds such as oxybenzone, avobenzone, octocrylene, and octinoxate have shown to reduce photosynthetic rates and increase the presence of reactive oxygen species that take DNA<sup>73</sup>. Thus, it may be beneficial to reduce the amount of SPF products that contain these substances and rather use approved ingredients such as titanium oxide and zinc oxide.

## VII. References

1. ACS Medical Content and News Staff. (2022, January 12). Risk of Dying from Cancer Continues to Drop at an Accelerated Pace. Retrieved August 3, 2023, from  
<https://www.cancer.org/research/acs-research-news/facts-and-figures-2022.html>
2. Stern, R. S. (2010). Prevalence of a history of skin cancer in 2007. Archives of Dermatology, 146(3). <https://doi.org/10.1001/archdermatol.2010.4>
3. Guy Jr, G. P., Thomas, C. C., Thompson, T., Watson, M., Massetti, G. M., & Richardson, L. C. (n.d.). Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. National Center for Biotechnology Information.  
<https://pubmed.ncbi.nlm.nih.gov/26042651/>
4. American Cancer Society. (2023). Melanoma of the skin. Retrieved August 3, 2023, from  
<https://cancerstatisticscenter.cancer.org/#!/cancer-site/Melanoma%20of%20the%20skin>
5. American Cancer Society. (2019, July 10). Ultraviolet (UV) Radiation. Retrieved August 3, 2023, from <https://www.cancer.org/cancer/risk-prevention/sun-and-uv/uv-radiation.html>
6. The Sun, UV, and You: A Guide to SunWise Behavior. (n.d.).  
<https://www.epa.gov/sites/default/files/documents/sunuvu.pdf>
7. Environmental Protection Agency. (2023, July 6). Ultraviolet (UV) Radiation and Sun Exposure. Retrieved August 3, 2023, from  
<https://www.epa.gov/radtown/ultraviolet-uv-radiation-and-sun-exposure>
8. Pérez-sánchez, A., Barrajón-catalán, E., Herranz-lópez, M., & Micol, V. (2018). Nutraceuticals for skin care: A comprehensive review of human clinical studies. Nutrients, 10(4), 403. <https://doi.org/10.3390/nu10040403>
9. FDA. (2020, August 19). Ultraviolet (UV) Radiation. Retrieved August 3, 2023, from  
<https://www.fda.gov/radiation-emitting-products/tanning/ultraviolet-uv-radiation>
10. Sinha, R. P., & Häder, D.-P. (2002). UV-induced DNA damage and repair: A review. Photochemical & Photobiological Sciences, 1(4), 225-236.  
<https://doi.org/10.1039/b201230h>
11. National Center for Environmental Health. (2023, July 3). UV Radiation. Retrieved August 4,

2023, from <https://www.cdc.gov/nceh/features/uv-radiation-safety/index.html>

12. Environmental Protection Agency. (2021, October 7). Basic Ozone Layer Science. EPA. <https://www.epa.gov/ozone-layer-protection/basic-ozone-layer-science>
13. NASA. (n.d.). World of change: Antarctic ozone hole. NASA. <https://earthobservatory.nasa.gov/world-of-change/Ozone>
14. Wendel, J. (2022, October 22). Hole in the ozone layer has grown for a 3rd year in a row - but scientists aren't concerned. Space.com. <https://www.space.com/earths-ozone-hole-biggest-since-2015-scientists-arent-concerned>
15. Ramsayer, K. (2022, October 25). Ozone Hole continues shrinking in 2022, NASA and NOAA scientists say. NASA. <https://www.nasa.gov/esnt/2022/ozone-hole-continues-shrinking-in-2022-nasa-and-noaa-scientists-say>
16. Merin, K. A., Shaji, M., & Kameswaran, R. (2022). A review on sun exposure and skin diseases. Indian journal of dermatology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9971785/>
17. Pandel, R., Poljšak, B., Godic, A., & Dahmane, R. (2013, September 12). Skin photoaging and the role of antioxidants in its prevention. ISRN dermatology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3789494/>
18. Hart, P. H., & Gorman, S. (2013, February). Exposure to UV wavelengths in sunlight suppresses immunity. to what extent is UV-induced vitamin D3 the mediator responsible?. The Clinical biochemist. Reviews. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3626364/>
19. Katiyar, S. K. (2007, September 18). UV-induced immune suppression and photocarcinogenesis: Chemoprevention by Dietary Botanical Agents. Cancer letters. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1995595/>
20. Juzeniene, A., & Moan, J. (2012, April 1). Beneficial effects of UV radiation other than via vitamin D production. Dermato-endocrinology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3427189/>
21. Brenner, M., & Hearing, V. J. (2008). The protective role of melanin against UV damage in human skin. Photochemistry and photobiology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2671032/>
22. Nakai, K., & Tsuruta, D. (2021, October 6). What are reactive oxygen species, free radicals, and oxidative stress in skin diseases?. International journal of molecular sciences. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8509443/>
23. Melanoma of the skin - cancer stat facts. SEER. (n.d.). <https://seer.cancer.gov/statfacts/html/melan.html>
24. Skin cancer facts & statistics. The Skin Cancer Foundation. (2023, March 6). <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>
25. Signs of melanoma skin cancer: Symptoms of melanoma. Signs of Melanoma Skin Cancer | Symptoms of Melanoma | American Cancer Society. (n.d.). <https://www.cancer.org/cancer/types/melanoma-skin-cancer/detection-diagnosis-staging/signs-and-symptoms.html>

26. Melanoma staging. Roswell Park Comprehensive Cancer Center. (n.d.).  
<https://www.roswellpark.org/cancer/melanoma/diagnosis/staging>
27. Melanoma survival rates: Melanoma survival statistics. Melanoma Survival Rates | Melanoma Survival Statistics | American Cancer Society. (n.d.).  
<https://www.cancer.org/cancer/types/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html>
28. Keung, E. Z., & Gershenwald, J. E. (2018, August). The Eighth edition American Joint Committee on cancer (AJCC) melanoma staging system: Implications for melanoma treatment and care. Expert review of anticancer therapy. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7652033/>
29. Ansstas, G., Poole, C. M., & Kottschade, L. A. (n.d.). The Journey Through Stage III Melanoma: A Guide for Patients. Conquer: the journey informed. <https://conquer-magazine.com/issues/special-issues/the-journey-through-stage-iii-melanoma-a-guide-for-patients>
30. Li, J., Uchida, T., Todo, T., & Kitagawa, T. (2006). Similarities and differences between cyclobutane pyrimidine dimer photolyase and (6-4) photolyase as revealed by resonance raman spectroscopy. Journal of Biological Chemistry, 281(35), 25551-25559.  
<https://doi.org/10.1074/jbc.m604483200>
31. Daya-grosjean, L., & Sarasin, A. (2000). UV-specific mutations of the human patched gene in basal cell carcinomas from normal individuals and xeroderma pigmentosum patients. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 450(1-2), 193-199.  
[https://doi.org/10.1016/s0027-5107\(00\)00025-7](https://doi.org/10.1016/s0027-5107(00)00025-7)
32. Maner, B. S., Dupuis, L., Su, A., Jueng, J. J., Harding, T. P., Meisenheimer, J., Siddiqui, F. S., Hardack, M. R., Aneja, S., & Solomon, J. A. (2020). Overview of genetic signaling pathway interactions within cutaneous malignancies. Journal of Cancer Metastasis and Treatment, 2020.  
<https://doi.org/10.20517/2394-4722.2020.60>
33. John Hopkins Medicine. (n.d.). BRAF Mutation and Cancer. Retrieved August 3, 2023, from <https://www.hopkinsmedicine.org/health/conditions-and-diseases/braf-mutation-and-cancer>
34. Ikehata, H., & Ono, T. (2011). The mechanisms of UV mutagenesis. Journal of Radiation Research, 52(2), 115-125. <https://doi.org/10.1269/jrr.10175>
35. Brash, D. E. (2014). UV signature mutations. Photochemistry and Photobiology, 91(1), 15-26.  
<https://doi.org/10.1111/php.12377>
36. McCain, J. (n.d.). he MAPK (ERK) Pathway: Investigational Combinations for the Treatment Of BRAF-Mutated Metastatic Melanoma. National Center for Biotechnology Information.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628180/>
37. National Center for Biotechnology Information. (n.d.). <https://www.ncbi.nlm.nih.gov/gene/5727>
38. National Human Genome Research Institute (Ed.). (2023, August 2). Cell Cycle. Retrieved August 3, 2023, from <https://www.genome.gov/genetics-glossary/Cell-Cycle#:~:text=A%20cell%20cycle%20is%20a,mitosis%2C%20and%20completes%20its%20division>

39. Khan Academy . (n.d.). DNA Excision Repair . DNA proofreading and repair. Retrieved July 29, 2023, from <https://www.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-replication/a/dna-proofreading-and-repair>.
40. Swift, L.h., & Golsteyn, R.m. (2016). The relationship between checkpoint adaptation and mitotic catastrophe in genomic changes in cancer cells. *Genome Stability*, 373-389. <https://doi.org/10.1016/b978-0-12-803309-8.00022-7>
41. professional, C. C. medical. (n.d.). Melanin: What is it, Types & Benefits. Cleveland Clinic. <https://my.clevelandclinic.org/health/body/22615-melanin>
42. VJ; , B. M. (n.d.). *The protective role of melanin against UV damage in human skin.* Photochemistry and photobiology. <https://pubmed.ncbi.nlm.nih.gov/18435612/>
43. National Human Genome Research Institute (Ed.). (2023, August 2). Cell Cycle. Retrieved August 3, 2023, from <https://www.genome.gov/genetics-glossary/Cell-Cycle#:~:text=A%20cell%20cycle%20is%20a,mitosis%2C%20and%20completes%20its%20division>
44. Davis, L. G., & Battey, J. F. (1986). Restriction Endonucleases (REs) and Their Use. Retrieved August 3, 2023, from <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/restriction-endonucleases>
45. Mandal, A. (2022, December 30). What is DNA Polymerase? Retrieved August 3, 2023, from <https://www.news-medical.net/life-sciences/What-is-DNA-Polymerase.aspx#:~:text=DNA%20polymerases%20are%20enzymes%20that,from%20one%20original%20DNA%20molecule>
46. Yousef, H., Alhajj , M., & Sharma, S. (2022, November 14). *Anatomy, skin (integument), epidermis - StatPearls - NCBI Bookshelf.* Anatomy, Skin (Integument), Epidermis. <https://www.ncbi.nlm.nih.gov/books/NBK470464/>
47. Tomkinson, A. E., & Sallmyr, A. (2013). Structure and function of the DNA ligases encoded by the mammalian lig3 gene. *Gene*, 531(2), 150-157. <https://doi.org/10.1016/j.gene.2013.08.061>
48. Skoczyńska, A., Budzisz, E., Trznadel-Grodzka, E., & Rotsztejn, H. (2017, April 13). *Melanin and lipofuscin as hallmarks of Skin aging.* Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii. <https://doi.org/10.5114/ada.2017.67070>
49. Sandnes Skin Clinic. (2019, May 20). How to choose the right sunscreen. Retrieved August 4, 2023, from <https://sandneshudklinikk.no/2019/05/20/how-to-choose-the-right-sunscreen/>
50. JM; , A. R. (n.d.). *Oxygen toxicity and reactive oxygen species: The Devil is in the details.* Pediatric research. <https://pubmed.ncbi.nlm.nih.gov/19390491/>
51. American Cancer Society. (2019, July 23). How Do I Protect Myself from Ultraviolet (UV) Rays? Retrieved August 3, 2023, from <https://www.cancer.org/cancer/risk-prevention/sun-and-uv/uv-protection.html>

52. Food & Drug Administration. (2023, May 16). Facts About Sunscreen. Retrieved August 3, 2023, from <https://www.fda.gov/news-events/rumor-control/facts-about-sunscreen>
53. American Academy of Dermatology Association (Ed.). (2023, July 19). Sunscreen FAQs. Retrieved August 3, 2023, from <https://www.aad.org/media/stats-sunscreen#:~:text=Dermatologists%20recommend%20using%20a%20sunscreen,of%20the%20sun%27s%20UVB%20rays>
54. Food & Drug Administration. (2023, May 24). Sunscreen: How to Help Protect Your Skin from the Sun. Retrieved August 3, 2023, from <https://www.fda.gov/drugs/understanding-over-counter-medicines/sunscreen-how-help-protect-your-skin-sun>
55. Harvard Health Publishing. (2017, June 1). Which is best for optimal sun protection — sprays or lotions? Retrieved August 4, 2023, from <https://www.health.harvard.edu/skin-cancer/which-is-best-for-optimal-sun-protection-sprays-or-lotions>
56. Smijs, T., & Pavel. (2011). Titanium dioxide and zinc oxide nanoparticles in sunscreens: Focus on their safety and effectiveness. *Nanotechnology, Science and Applications*, 95. <https://doi.org/10.2147/nsa.s19419>
57. Couteau, C., Alami, S., Guitton, M., Paparis, E., & Coiffard, L.J. (n.d.). Mineral filters in sunscreen products--comparison of the efficacy of zinc oxide and titanium dioxide by in vitro method. National Center for Biotechnology Information. <https://pubmed.ncbi.nlm.nih.gov/18271305/>
58. Harvard Health Publishing. (2021, February 15). The science of sunscreen. Retrieved August 3, 2023, from <https://www.health.harvard.edu/staying-healthy/the-science-of-sunscreen>
59. Food & Drug Administration. (2020, January 21). Shedding More Light on Sunscreen Absorption. Retrieved August 3, 2023, from <https://www.fda.gov/news-events/fda-voices/shedding-more-light-sunscreen-absorption>
60. Suh, S., Pham, C., Smith, J., & Mesinkovska, N. A. (2020). The banned sunscreen ingredients and their impact on human health: A systematic review. *International Journal of Dermatology*, 59(9), 1033-1042. <https://doi.org/10.1111/ijd.14824>
61. Inselman, A. L. (2016, June 23). Effect of Oxybenzone on Fertility and Early Embryonic Development in Sprague-Dawley Rats (Segment I). Retrieved August 3, 2023, from [https://ntp.niehs.nih.gov/sites/default/files/nctr/e0218601\\_report\\_508.pdf](https://ntp.niehs.nih.gov/sites/default/files/nctr/e0218601_report_508.pdf)
62. Mohammadi, S., Saberidokht, B., Subramaniam, S., & Grama, A. (2015). Scope and limitations of yeast as a model organism for studying human tissue-specific pathways. *BMC Systems Biology*, 9(1). <https://doi.org/10.1186/s12918-015-0253-0>
63. An Introduction to *Saccharomyces cerevisiae*. (2023). Retrieved August 3, 2023, from <https://app.jove.com/v/5081/an-introduction-to-emsaccharomyces-cerevisiaeem>
64. Botstein, D., Chervitz, S. A., & Cherry, M. (1997). Yeast as a model organism. *Science*, 277(5330), 1259-1260. <https://doi.org/10.1126/science.277.5330.1259>
65. Scharer, O. D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(10), a012609. <https://doi.org/10.1101/cshperspect.a012609>

66. SuiGenerisBrewing. (2013, June 21). First wild yeast hunt. Retrieved August 4, 2023, from <https://suigenerisbrewing.com/index.php/2013/06/21/first-wild-yeast-hunt/>
67. *Bacteria Growth Curve*. (2022). OD600 (Cell Density, Bacterial Growth, Yeast Growth). Retrieved August 3, 2023, from <https://www.implen.de/od600-diluphotometer/od600/>.
68. Venosa, A. (2017, August 11). Dress to Protect: 5 Things that Affect How Well Your Clothes Block UV Rays. Retrieved August 4, 2023, from <https://www.skincancer.org/blog/dress-to-protect-5-things-that-affect-how-well-your-clothes-block-uv-rays/#:~:text=Dark%20or%20bright%20colors%2C%20including,protection%20the%20clothing%20will%20provide>
69. The trouble with ingredients in sunscreens: EWG's guide to sunscreens. The trouble with ingredients in sunscreens | EWG's Guide to Sunscreens. (n.d.).  
<https://www.ewg.org/sunscreen/report/the-trouble-with-sunscreen-chemicals/>
70. Reflection of ultraviolet light. (n.d.). [https://library.med.utah.edu/kw/derm/pages/meet\\_7.htm](https://library.med.utah.edu/kw/derm/pages/meet_7.htm)
71. Heerfordt, I. M., Torsnes, L. R., Philipsen, P. A., & Wulf, H. C. (2018, March 28). Sunscreen use optimized by two consecutive applications. PloS one.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5874020/>
72. Fiume, M., Haldreth, B., Bergfeld, W. F., Belsito, D. V., Hill, R. A., Klaassen, C. D., Lieber, D.C., Marks, J. G., Jr., Shank, R. C., Slaga, T. J., Snyder, P. W., & Andersen, F. A. (2015, December). Safety Assessment of Alkyl Ethylhexanoates as Used in Cosmetics. National Library of Medicine. Retrieved August 3, 2023, from [https://pubmed.ncbi.nlm.nih.gov/26684798/#:~:text=The%20Cosmetic%20Ingredient%20Review%20\(CIR,when%20formulated%20to%20be%20nonirritating.](https://pubmed.ncbi.nlm.nih.gov/26684798/#:~:text=The%20Cosmetic%20Ingredient%20Review%20(CIR,when%20formulated%20to%20be%20nonirritating.)
73. Zhong, X., Downs, C. A., Che, X., Zhang, Z., Li, Y., Lou, B., Li, Q., Li, Y., & Gao, H. (2019, November). The toxicological effects of oxybenzone, an active ingredient in suncream personal care products, on prokaryotic alga *Arthrospira* sp. and eukaryotic alga *Chlorella* sp. Aquatic Toxicology. Retrieved August 3, 2023, from <https://www.sciencedirect.com/science/article/pii/S0166445X19305934>

# **Investigating the Impacts of Environmental Factors on Neuron Function Using *L. terrestris* and *C. elegans***

Vanesa Aguay, Sanjana Balaji, Noah Breithaupt, Renuk DeAlmeida, Namya Jindal, Gayath Karunaratne, Sanjay Nair, Savannah Stackhouse, Diya Thomas, Wilson Zhang

## **Abstract**

This study aims to test the effects of various environmental factors on the neuron function of *L. terrestris* worms as a model organism to see their respective impacts on humans and on environmental health. We used *L. terrestris* (earthworms) as model organisms for neural activity and *C. elegans* as a model organism for behavior. The effects of different temperatures, pH levels, microplastic concentrations, road salt concentrations, and sugar beet additives to de-icers on the conduction velocities of earthworms were measured using action potential bioamplifiers, or SpikerBoxes. It was hypothesized that these environmental factors would all lead to neurodegeneration and decreased conduction velocity. The SpikerBox was utilized to record action potentials within the central nerve cord of *L. terrestris*, which was then used to calculate conduction velocity. Our results show that the environmental factors do affect neural function, but in some unexpected ways.

## **I. Introduction**

Human industrial activity is the principal source of environmental pollution. The release of sulfur and nitrogen oxides into the atmosphere by fossil fuels and vehicular exhaust has resulted in the common occurrence of acid rain, harming ecosystems, and heightening global warming. Road salts have been used to prevent the formation of ice on roads. However, little research has been done on the effect of sugar beet additives in mitigating the negative ramifications of road salt runoff. As a result of the skyrocketing demand for plastic, microplastic pollution is at an all-time high and the effects of its substituents are relatively unknown. All these pollutants can have adverse effects on humans' neurological health by inducing neurodegeneration, which is the process by which neurons progressively lose their capability to fire and trigger responses. These pollutants can also affect the neural functioning of species in the ecosystem. By using worms as a model species, we can observe these environmental effects on neurological function and assess both human health and environmental health.

### **a. A Gap in Literature**

Since the rise of environmental advocacy, there has been an increased awareness of industrial pollutants. In fact, a study performed by the Columbia Climate School revealed that 76% of Americans would describe themselves as "environmentalists," compared to 56% in 1989.<sup>1</sup> With negative effects of pollution continuing to persist and increased public awareness, it has become clear that something must be done.

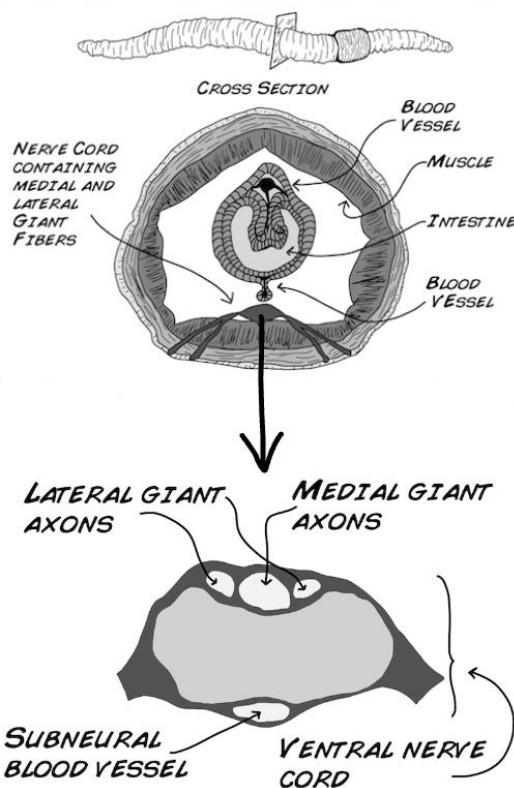
Recognizing this need, scientists have continued to pursue research in the environmental sciences on how pollutants, temperature fluctuations, changes in pH and other variables can affect the environment.<sup>2</sup> However due to the ever-changing environmental conditions and the incessant use of new chemicals, much of this research finds itself outdated: not because of time elapsed, but because of changing conditions and practices. Due to these circumstances, studies performed merely years ago in different conditions feed uninformed policy and an inaccurate view of environmental pollutants,<sup>3</sup> thus establishing the importance of

research continuity of possible environmental dangers, namely climate change and use of chemicals. While a gap in the literature can be easily observed in all six testing environments, the study as a whole will aid in closing the gap in modern environmental science research by investigating both the causes of environmental pollutants, namely road salts and microplastics, along with the effects of such pollutants, specifically climate change and pH fluctuations. To do so, this study employs the uses of model species, *C. elegans* and *L. terrestris*.

### b. The Use of Model Species

The study focused on the *Lumbricus terrestris*, or the earthworm, due to its significant role as both a bioindicator of ecosystems and as a model organism. *L. terrestris* serves as a keystone species due to its role as an ecosystem engineer where in many terrestrial ecosystems, they constitute 40%-90% of soil macrofauna biomass<sup>4</sup>. In addition to frequently affecting soil temperatures as the foundation of many ecosystems, they serve as indicators of the health of the ecosystem which serves as the first reason for the use of this subject. Ecosystems are also susceptible to environmental changes, and these impacts, including those due to human activity, have a potent effect on organisms within. These activities have recently been speculated to have an effect on the function of organisms' nervous systems such as the potential linkage of warmer climates to neurodegenerative disorders.<sup>5</sup> Humans, however, have over 100 billion neurons and 100 trillion synaptic connections, so due to the high complexity of nervous systems in mammals, it is easier to measure and study the nervous system in model organisms<sup>6</sup>. The second reason for choosing the earthworm in the study is due to its role as a model organism with its simple and integrated nervous system that conserves the same principles of conduction velocity and myelination as seen in mammalian nervous systems.

The nervous system of *L. terrestris* consists of the peripheral nerves, the cerebral ganglia serving as a simple brain, and the ventral nerve cord serving as the functional equivalent of a spine to a vertebrate. The cerebral ganglia is connected to a ventral nerve cord that runs the length of the body and which is responsive for coordinating neural signaling as sensory information is integrated to allow for the conduction of synaptic transmission. We will place the electrode pins near the ventral nerve cord to measure the action potential as it travels down the axon. Within the ventral nerve cord are the dorsal giant axons which contain two lateral giant fibers as well as one medial giant fiber that receives sensory information depending on which part of the worm is stimulated (Figure 1). When the anterior end of the worm is stimulated, the sensory information is transmitted through the medial giant fiber, while stimulation in the skin cells of the posterior end of the worm transmits sensory information through the lateral giant fibers (each of which has a smaller axon diameter than does the medial giant fiber).



**Figure 1: Anatomy of *L. terrestris* Ventral Nerve Cord<sup>7</sup>**

### c. Ethical Considerations

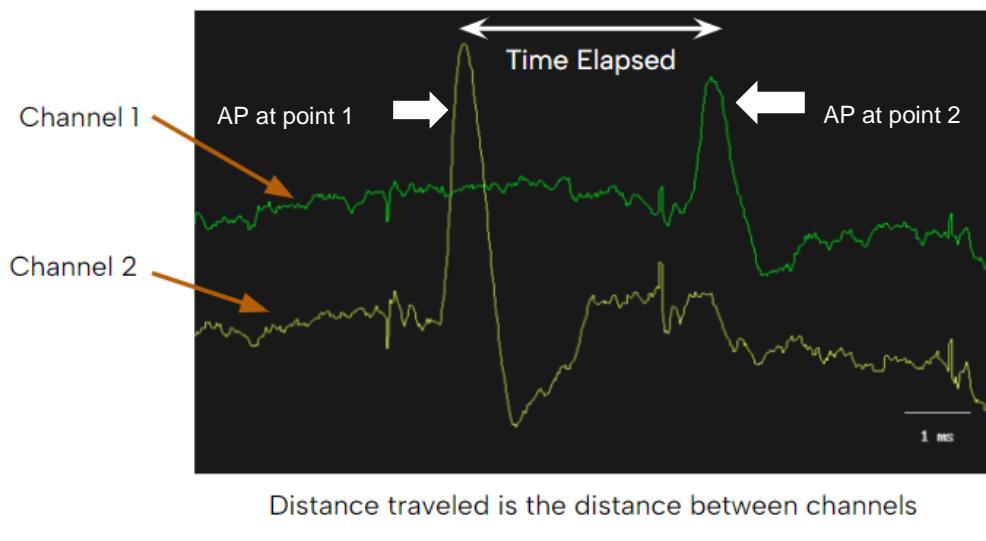
Model species have continuously been used for research for over 2,400 years.<sup>8</sup> However, many ethical considerations, including the realization that animals can feel pain and the fact that many experiments cause an unnecessary amount of suffering, have arisen. In regards to this, we first attempted the use of a noninvasive approach to recording action potentials detailed by Shannon et al<sup>11</sup> but were not able to record any true action potentials. This conclusion allowed our project to continue on by using invasive methods where the worm is pierced by electrodes in order to get stronger signals nearer to the lateral giant fiber. Our worms were utilized sparingly and humanely as we provided them with ample time to recover so fewer worms would be needed for our experiments. Worms were only replaced if absolutely necessary.

### d. The Biology of Action Potentials and Conduction Velocity

An action potential is an electrical neural signal, mediated by voltage-dependent changes in ionic permeability, that travels down the axon of a neuron. Action potentials induce the release of neurotransmitters, enabling communication between neurons, secretion of hormones, and stimulation of muscle cells. The action potential is driven by the ratio of intracellular and extracellular ions. Embedded in the cell membrane are voltage-gated (VG) sodium and potassium ion channels which can open and close, varying ionic permeabilities and creating different polarizations of the cell. The cell membrane is more permeable to K<sup>+</sup> than Na<sup>+</sup> due to potassium leak channels embedded in the membrane, and potassium ions and anionic proteins are more abundant intracellularly while sodium and chloride ions are present in higher concentrations extracellularly. These factors cause the cell membrane to have an approximate voltage of -70 mV at rest. The action potential is the cascade of electrical activity that is created upon

depolarizing stimulus. The incoming depolarizing stimuli must meet the -55 mV threshold potential in order for an action potential to be generated. When this happens, VG sodium channels open, which allow  $\text{Na}^+$  to flood into the cell. The influx of positive charge inside the cell causes the voltage difference between the inside and outside of the neuron to decrease. Further depolarization triggers the opening of additional VG sodium channels, increasing  $\text{Na}^+$  permeability. As the intracellular concentration of  $\text{Na}^+$  increases, membrane potential shifts in favor of the sodium's equilibrium potential of 55 mV. The action potential reaches a peak at approximately 35 mV and at this point, the sodium channels inactivate. Simultaneously, the VG potassium channels open, and repolarization occurs. Potassium ions flow out of the cell, driving the membrane potential towards potassium's equilibrium potential of -90 mV. As a result, the cell's membrane potential drops to about -75mV and the cell becomes hyperpolarized. The VG potassium ion channels close and the sodium potassium pump maintains the normal concentrations of  $\text{Na}^+$  and  $\text{K}^+$  removing 3  $\text{Na}^+$  ions from inside the cell and bringing in 2  $\text{K}^+$  ions into the cell.<sup>9 10</sup>

Conduction velocity (CV) is a measure of how fast electrical signals move through a nerve. This is done by measuring the times at which the action potential is recorded at different positions along the worm. By measuring the distance between electrodes and dividing by the time it takes for the action potential to be recorded at different locations along the ventral nerve cord (Figure 2), we can determine the velocity at which the signal travels through the organism. In this study, we tested the effects of different treatments on the worms by measuring their conduction velocities under different conditions.



**Figure 2: Action potential as Recorded in Spike Recorder**

$$\text{Conduction Velocity} = \frac{\text{distance traveled}}{\text{time elapsed}}$$

### e. Collection of Data

In this experiment, a Neuron SpikerBox, a bioamplifier that allows you to both hear and see action potentials, will be used to record the electrochemical signals in the giant nerve fibers of the worms. Specifically in earthworms, there are three primary nerve fibers: one Medial Giant Nerve Fiber (MGF) and two Lateral Giant Nerve Fibers (LGF). These fibers carry electrical signals (action potentials) down the

length of the worm, which allows the worm to react to different stimuli. The lateral giant nerve fiber is activated when the posterior end (back) of the worm is stimulated while the medial giant fiber is measured when the anterior (front) end of the worm is stimulated.<sup>11</sup> Measurements were taken from the lateral giant fiber (LGF) to find the time between peaks and the conduction velocities. Using the data collected from the action potentials, the conduction velocity can be calculated. This study will be observing the effect of numerous environmental factors on the action potential and conduction velocity of *L. terrestris*. Additionally, this study will be testing the effects of some of these environmental factors on the behavior of *C. elegans*, a microscopic worm.

#### **f. Temperature**

According to the Climate Science Special Report, Earth's temperature has been rising by 0.14 degrees Fahrenheit on average per decade for a total of 2 degrees higher than the pre-industrial average.<sup>12</sup> Although natural variability does play a role, there is an irrefutable link between the increasing use of fossil fuel burning and increasing temperatures, which have altered the global climate and therefore impacted organisms across the world. For example, in an article by Seung-Jae Lee et al, it was indicated that temperature could impact earthworms' thermosensory neurons to the extent that their rate of aging was changed.<sup>13</sup> By studying the effects of temperature on earthworms' central nervous systems, our understanding of how the changing climate affects organisms' neurological processes can be improved.

In this experiment, groups of *L. terrestris* were kept in different simulated climates. Ideally, very small increments would have been used between different temperature groups so as to mirror the kinds of incremental increases we observe in the real climate. However, this would be unfeasible given the time and resource constraints. As such, 5 temperature conditions were tested starting with the control group being at 10 degrees Celsius, since that is the temperature *L. terrestris* naturally experiences,<sup>14</sup> and each subsequent group's temperature increased by an increment of 5 degrees Celsius. Sampling these conditions will still provide ample data to observe the effects of climate change on the neurological functions of *L. terrestris* as they are representative of future warmer climate-change-induced ecosystems.

#### **g. pH**

One of the factors this study will test is how different pH levels, especially ones outside a worm's optimal range, will affect the neuron function of an earthworm. In the environment, pH changes occur due to events such as acid rain, pollution, or temperature<sup>15</sup>. Earthworms often thrive at a neutral pH, similar to humans; however, they have a higher tolerance for pH variations, ranging from 5.0 to 8.0. In contrast, humans must stay between a pH range of 7.35 to 7.45 to maintain homeostasis. Acidosis and alkalosis are two conditions resulting from unbalanced pH levels in the body. Acidosis is when there is too much acid in the body fluids, whereas alkalosis is when there is too much base in the body fluids.<sup>16</sup> In a recent study, it was found that pH values which ranged from 5.1 - 8.6 resulted in earthworm death; however, there is no relevant study to prove how pH outside of an organism's tolerance will affect neuron function.<sup>17</sup> The pH of the soil will be adjusted by using soil-specific acid and base solutions. Based on the study, we hypothesize that a change in pH will cause the rate of action potentials to decrease which will decrease conduction velocity.

#### **h. Phthalate Alternative- ATBC**

Microplastics are now found in our food, bodies, and environment, and understanding the effects of microplastic pollution on human health is imperative to ensuring the safety of the general public in regards to plastic pollution. Microplastics make their way into soil through sewage sludge, which is often used as

fertilizers. Sewage sludge is the by-product from sewage treatment of industrial or municipal wastewater. In our study, we were most interested in phthalates, which are plasticizers used in almost every plastic product to enhance their durability and elasticity. However, they have shown damaging effects to the endocrine and reproductive systems in a variety of model organisms.<sup>18</sup> Phthalate exposure has also caused breaks in the DNA of stem cell derived human neurons<sup>19</sup>. To combat the side effects caused by phthalates, companies introduced acetyl tributyl citrate (ATBC) as an alternative since it has been heralded as a safer alternative, but it has not been thoroughly studied and its neural effects have not yet been quantified.<sup>19</sup> We plan on testing the effects of ATBC on neurodegeneration in a generation of earthworms by examining neural firing throughout the lifespan. We hypothesize that the ATBC will cause neurological damage because previous studies suggest it could have adverse effects on the generation of glial cells<sup>20</sup>. This can cause accelerated and permanent neurodegeneration which would slow neural signaling. It is because of this that we expect to see a decrease in the conduction velocity of worms with more exposure to ATBC.

#### i. Road Salts

In many countries across the world when it starts to snow, the local municipalities start to place salts on the road to prevent people and cars from slipping on frozen surfaces. After the snow melts from the salts, the salts are carried away into the ground. The salts in the ground then can be digested or absorbed by the various organisms living in the soil, such as earthworms. To determine the impact of this, we plan to investigate the effects of road salts on *L. terrestris*. From our research, we found that some states use salt brine to ice roads which contain around 23.3% of salt.<sup>21</sup> For our experiment, we dissolved different concentrations of CaCl<sub>2</sub> and KCl in water. The solutions were then poured into the soil of different groups of worms to simulate road salt treatment.

We predicted that the ions contained in the salts would have varying effects on the neurons of the earthworms. KCl may overload the resting potassium and chloride pumps and make the cell have a greater negative charge. This will increase the reversal potential because the flow will change from negative to positive. The increased reversal potential will decrease the frequency of action potentials in the worms. The lower frequency of action potentials will cause sluggish movements and paralysis in the earthworms. We predicted that the increase in CaCl<sub>2</sub> will cause a calcium rush in the synaptic terminals which will release neurotransmitters and use vesicles. That vesicle use will be inefficient as it will have to compensate for the increased calcium chloride leading to the neuron eventually running out of vesicles and reduce the ability of the presynaptic neuron to act on the postsynaptic target. In an excitatory circuit, this could lead to a decreased likelihood of action potential firing.

#### j. Sugar Beet Juice Additives in Road Salt Brine

Sodium Chloride (NaCl) has historically been one of the most utilized methods of de-icing roads due to its property of freezing point depression.<sup>22</sup> However, many have begun to criticize such methods of de-icing, due to the excessive damage the chemical has caused. In fact, the United States Environmental Protection Agency (EPA), a highly influential government agency promoting environmental conservation, has continued to publicly denounce the excessive use of such chemicals, arguing that the current composition of salt brine can “contaminate drinking water, kill or endanger wildlife, increase soil erosion, and damage private and public property”.<sup>23</sup> Many plausible solutions have been proposed, including the use of sand, permeable roads, and the addition of different chemicals. However, these salt brine alternatives have proven to be costly and some, namely sand, cause unwarranted environmental impacts.<sup>24</sup> Because of this cost and the lingering possibility of further environmental damage, many communities are not willing to actively pursue these possible resolutions.

As numerous solutions continue to prove impractical and communities continue to suffer due to the effects of salt brine, environmentalists have continued to seek ways to mitigate the negative effects of the de-icer, while maintaining the positive ones, all while avoiding large expenses. Many areas have already employed a pragmatic solution: the unlikely use of sugar beets.

States such as Missouri started using sugar beet additives in 2006<sup>25</sup> and have seen success in using the additives<sup>26</sup> while other areas, such as Michigan have passed bills requiring the state's department of transportation to test the addition of organic additives to both limit the amount of road salt and have the solution better stick to the road surface. These additives mainly consist of Sugar Beet juice.

It is because of this increased use of sugar beet-based additives, and the ongoing damage of the current popular salt brine usage on roads, specifically on the environment, that this portion of the portion of the study aims to answer the question, "How do normal salt brine and salt brine with a Sugar Beet additive affect the neurons of model organisms *Caenorhabditis elegans* and *Lumbricus terrestris*?" To answer this, we hypothesize that application of both treatments containing salt brine will lower the conduction velocity. However, the de-icer including the sugar beet additive will prove to be more similar to the control.

#### **k. Goal of Research**

This study aims to determine if industrial pollution affects the neurological function of model species *Lumbricus terrestris* and *Caenorhabditis elegans*. To accomplish this, we studied the effects of manmade environmental changes on the neural function of different earthworm species via measurement of the conduction velocity. In this study, we show the impacts of plasticizers, pH, road salts, sugar beet additives, and temperature on the neuron function of *L. terrestris*.

## **II. Methodology**

For this study, approximately sixty adult *L. terrestris* worms were observed. The worms used for this study were Canadian nightcrawlers sold by the DMF Bait Company. For a majority of the treatments, a ratio of 80g of soil to 40g of water was used for the environment of the worms. The primary soil that was used for experimentation is Scott's topsoil. Tap water (ionized) was used rather than deionized water to prevent interference with the neuron function of the worms. The ions in tap water play a role in the electrochemical gradient in neurons which helps them conduct electrochemical signals. All the treatments were chronic, meaning the worms were exposed to treatments for 2-4 days before the conduction velocity was recorded. During experimentation the worms were stored in paper or styrofoam cups in a 4°C refrigerator because that is the optimal temperature for worms. For specific studies, this maintenance was adjusted to fit the treatment being observed.

To record the neurological function of the worms, action potentials were measured, as adapted from the American Physiological Society journal.<sup>11</sup> To measure the action potentials, SpikerBoxes (Backyard Brains) were used. The *L. terrestris* underwent an anesthetization procedure in order to conduct the following method of measuring conduction velocity. The worm was placed in 100 mL of 10% EtOH for 5 to 10 minutes until movement stopped. The worms were anesthetized to ensure that only neural activity is measured. To record an action potential, the anesthetized worms were pinned to a corkboard using 3 electrodes. The first pin, the black one, was a ground electrode that provides a baseline to which the other pins can be related. The second and third pins, red and white, were used to measure the action potentials of the LGF. The pins are positioned so the black pin is the closest to the clitellum and furthest away from the side that is stimulated, and then the red and white can be positioned interchangeably closer to the stimulated posterior

side. The pins were inserted at an angle next to the blood vessel into the dorsal side of the worm so that it would pass through the ventral nerve cord. To determine the conduction velocity, the worms were stimulated on the anterior or posterior side to stimulate the MGF or LGF, respectively. The program we used to record the action potential was Spike Recorder (Backyard Brains) which produced a graph of the neural signal. From the graphs, the difference in times between action potential peaks can be determined.

### a. Temperature

To precisely measure the effects of temperature on the conduction velocity of *L. terrestris*, five groups of worms were held in different temperature-controlled environments at 10, 15, 20, 25, and 30°C over a period of 72–96 hours prior to recording (Table 1). Confounding variables such as the soil composition, moisture levels, and number of worms per group were eliminated by keeping conditions constant across all groups: 80 grams of the same soil type, 30 milliliters of water, and 2 worms per group. The only exception to these homologized conditions was in the group held at 30 degrees Celsius where only one worm was used to minimize unnecessary harm to the worms as such conditions are likely fatal. Once removed from the temperature-controlled environments, each group of worms was placed into a bath of 10% ethanol solution and then water pre heated or cooled to the same temperature at which they were housed to anesthetize them. Once transferred to the SpikerBox for action potential recording, a consistent temperature was maintained by placing an ice bath for the cooler-than-room-temperature groups or a heat plate for the warmer-than-room-temperature groups underneath the measuring contraption. Maintaining a constant temperature immediately prior to and throughout the recording process allowed us to ensure that a sudden change in temperature would not impact any results from the experiment.

**Table 1: Temperature Experimental Groups and Treatments**

Temperature Group	Subjects	Treatment
Group 1: Natural Temperature (Control)	Worm 1 and Worm 2	10°C
Group 2: Slightly Elevated Temperature	Worm 3 and Worm 4	15°C
Group 3: Elevated Temperature	Worm 5 and Worm 6	20°C
Group 4: Highly Elevated Temperature	Worm 7 and Worm 8	25°C
Group 5: Extreme Temperature	Worm 9	30°C

### b. pH

The effect of pH on neuron function was tested by placing worms in 3 different environments: a control (neutral pH), alkaline (pH > 8.0) , and acidic (pH < 5.0) environment. The soil used was Scott's topsoil to replicate the environment that earthworms would normally inhabit. The pH of the soil was tested beforehand and was neutral for worms (7.5). The method used to calculate the pH of the soil before and after treatment was adapted from the USDA method (Soil Health - pH). <sup>27</sup> This method involved placing 30mL of soil and

30mL of water (1:1 ratio) into a test tube and shaking 25 times. The tube was left to sit for a minute, and a cap of soil was then poured for testing. This solution was again left to sit for an additional 2-3 minutes, and pH strips were used to find the appropriate pH levels of each sample. To provide a stable environment for the worms, 4 inches (approximately 80g) of soil was placed in paper cups. We used pH soil adjusters (pH Up and Down - Concentrated pH Adjuster for Optimal Soil and Feed Water by Bloom City) to increase or decrease the pH of the soil because soil has a high buffering capacity, and these solutions were soil-specific.<sup>28</sup> We created 2 diluted solutions based on the suggested measurement and according to the following calculation:

$$\frac{1 \text{ gal } H_2O}{3785 \text{ mL adjuster}} = \frac{x}{500 \text{ mL adjuster}} = 0.132 \text{ gal } H_2O$$

Based on this calculation and the amount of soil that was added, we decided to mix 20mL of the acid or base solution into the soil of each of the experimental groups (acid solution to pH < 5.0 group and base solution to pH > 8.0 group). Each of the groups also had water mixed into the soil to maintain a moist environment: control group - 40 mL, acid/base groups - 20 mL. After the pH of soil was adjusted, we placed 2 worms in each group and allowed them to settle for 3 days before measuring action potentials.

### c. Phthalate Alternative - ATBC

To determine the concentration of Acetyl Tributyl Citrate (Sigma-Aldrich, 388378) to use in the soil, we found two main pieces of information: the concentration of plasticizer in common plastics, and the concentrations of these microplastics in various soil environments. We used PVC as our plastic of interest because it is a commonly found material in microplastic pollution. On the higher end, PVC has a 50% concentration of plasticizer by mass.<sup>29</sup> The average concentration of microplastics in municipal sites was 45 mg kg<sup>-1</sup> dry soil. In industrial sites, the concentrations rise 2 to 4 orders of magnitude (4.5 g kg<sup>-1</sup> dry soil - 450 g kg<sup>-1</sup> dry soil).<sup>30</sup>

With this information, we performed the following calculations to determine how much solution and water was needed, based on the fact that we were keeping our groups in 80g soil and we were adding our 97.5% pure ATBC to the soil, along with 40mL of ionized water.

#### Municipal Calculations:

$$\frac{45 \text{ mg plastic}}{1 \text{ kg soil}} * \frac{500 \text{ mg ATBC}}{1000 \text{ mg plastic}} * \frac{1 \text{ g}}{1000 \text{ mg}} * \frac{100 \text{ mL ATBC Solution}}{97.5 \text{ g ATBC}} * 0.08 \text{ kg soil} * \frac{1000 \mu\text{L}}{1 \text{ mL}} = \\ 1.846 \mu\text{L ATBC solution}$$

#### Mild Industrial Calculations (2 orders of magnitude)

$$\frac{4500 \text{ mg plastic}}{1 \text{ kg soil}} * \frac{500 \text{ mg ATBC}}{1000 \text{ mg plastic}} * \frac{1 \text{ g}}{1000 \text{ mg}} * \frac{100 \text{ mL ATBC Solution}}{97.5 \text{ g ATBC}} * 0.08 \text{ kg soil} * \frac{1000 \mu\text{L}}{1 \text{ mL}} = \\ 184.6 \mu\text{L ATBC solution}$$

Extreme Industrial Calculations (4 orders of magnitude)

$$\frac{450 \text{ g plastic}}{1 \text{ kg soil}} * \frac{0.5 \text{ g ATBC}}{1.0 \text{ g plastic}} * \frac{100 \text{ mL ATBC Solution}}{97.5 \text{ g ATBC}} * 0.08 \text{ kg soil} = 18.46 \text{ mL ATBC solution}$$

Each condition was accommodated with 40 mL ionized water. We used 2 worms per condition and were housed in treated soil for 3-4 days before testing.

**d. Road Salts**

We tested the effects of road salts on worms by splitting worms into three groups: one group exposed to KCl in water, one group exposed to CaCl<sub>2</sub> in water, and a control group exposed to only water. First, we determined the amount of water necessary to create a suitable environment for the worms, so we filled a paper cup with soil and poured water to make the soil damp. We found that around 88g of soil required 40mL of water. Next, we calculated the concentration of salt for our experimental groups with the salts KCl and CaCl<sub>2</sub>. To calculate the grams of salt required for a 26% concentration, we used the following equation:

$$40.0 \text{ grams water} * \frac{0.26 \text{ grams of salt}}{1 \text{ gram of water}} = 10.4 \text{ grams of salt}$$

We converted the grams of salt to molar mass with the following calculations:

$$10.4 \text{ grams CaCl}_2 * \frac{1 \text{ mol CaCl}_2}{110.98 \text{ grams of CaCl}_2} = 0.09 \text{ mol CaCl}_2$$

$$10.4 \text{ grams KCl} * \frac{1 \text{ mol KCl}}{74.55 \text{ grams of KCl}} = 0.14 \text{ mol KCl}$$

We then calculated the molar concentrations of the salts needed given the volume of 40 mL of water:

$$\frac{0.09 \text{ mol CaCl}_2}{0.04 \text{ L water}} = 2.25 \text{ M CaCl}_2$$

$$\frac{0.14 \text{ mol KCl}}{0.04 \text{ L water}} = 3.5 \text{ M KCl}$$

We used 2.25M solutions for our CaCl<sub>2</sub> as that was our calculated value. A concentration of 2.25M was also used for the KCl as that was the greatest concentration we could obtain. We poured the soil into the paper cup and used 40mL of tap water for the control group and 40mL of the salt solutions for the KCl and CaCl<sub>2</sub> groups. Two worms were then placed into each cup. The cups were then covered with plastic wrap with holes to allow oxygen to flow. Then they were stored at 4°C for 4 days. After our initial testing, we determined that the upper echelon of the percentage of saline solution was lethal to the worms, so we changed the solution to the minimum concentration of saline which is 3.5% for each salt. The resulting molarity was 0.5M of KCl and 0.3M of CaCl<sub>2</sub>. This concentration was also lethal for the worms, so the concentration of salt was reduced to 1.0%, which corresponds to molarities of 0.134M and 0.0901M for the KCl and CaCl<sub>2</sub> respectively. For quantitative data for our control group, we used values from published research as it would provide us with a more accurate comparison for our experimental groups.<sup>11</sup> We

recorded qualitative data observations for our control group to serve as a baseline to compare the other groups with.

### e. Sugar Beet Additive

#### 1. Concentrations of Salt and Beets

To conduct this experiment in the most accurate manner, concentrations of salt brine were based on the Missouri Department of Transportation's use of salt brine before the state started using the sugar beet juice additive (pre-2006) and after the state began using the additive (post-2006) due to the fact that the state has been using the additive for a time period greater than fifteen years and over a long span of area.<sup>25</sup> For the salt brine specifically, a ratio of 23.3% of sodium chloride by weight to 76.7% tap water was utilized. To create the sugar beet solution 80 percent of the lab manufactured salt brine and 20 percent of Graftschafter Goldschaft Original's Sugar Beet Molasses were mixed.

#### 2. Simulated Environment

Since soil pH and composition can affect neural activity and behavior in worm species, soil most comparable to the Missouri landscape was planned to be utilized. However, due to the lack of availability, along with time constraints, this was unable to be realized. A standard topsoil (Scott's Topsoil), bought from a hardware store was substituted.

To test which solution causes more runoff, three identical simulated roads were constructed, consisting of a bottom layer of soil, a layer of aggregate, and a top layer of soil, along with a 0.25 m x 0.15 m slab of asphalt (Figure 3 and Table 2). This material was held in three identical plastic containers for each separate group, measuring 0.3670 m x 0.3188 x 0.1440 m (length, width, depth respectively).

To maintain consistency, all three groups were held at the same temperature of four degrees Celsius and given the same amount of simulated precipitation every two days.

**Table 2: Construction of Simulated Environment**

	<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>
<b>Soil (Bottom Layer)</b>	550.62 g	550.31 g	550.24 g
<b>Aggregate</b>	1095.50 g	1095.43 g	1095.36 g
<b>Soil (Top Layer)</b>	847.73 g	847.87 g	847.90 g
<b>Asphalt</b>	0.155m x 0.25m	0.155m x 0.25m	0.155m x 0.25m



**Figure 3: Cross Sectional View into Layers of Artificial Road Environment**

### 3. Application of Treatment in *L. terrestris*

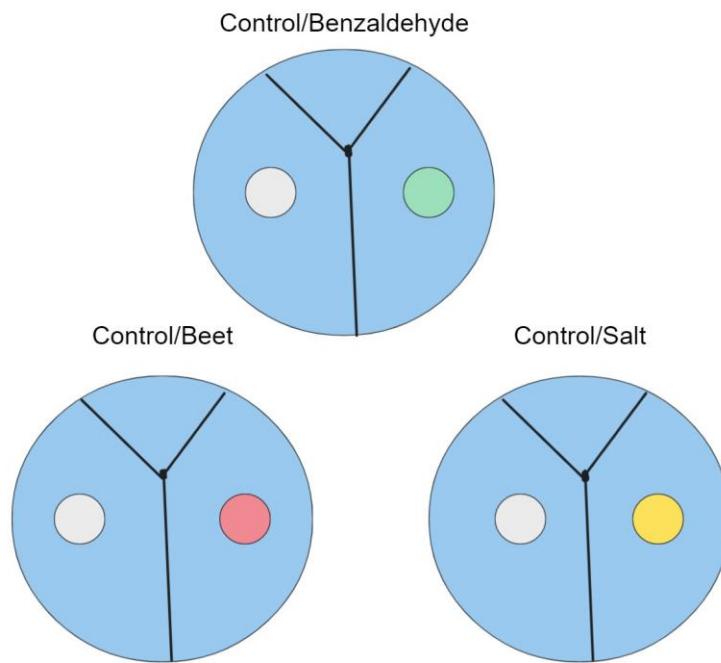
Throughout the duration of this experiment data was collected from three different groups: Control, Experimental A, and Experimental B. Treatments, as outlined in Table 3, were applied in a thin layer, in accordance with the methods that the Missouri Department of Transportation uses during the winter season.<sup>25</sup> After, in order to best simulate natural conditions, rain was simulated by spreading 2.73 inches of water across the container. This value is based on Missouri's average rainfall during the winter months. Worms were kept in the artificial road environments for four days at 4 °C before we recorded their action potentials.

**Table 3: Description of Treatment Groups**

Group of Worm	Treatment
Group 1: Control	N/A
Group 2: Experimental A	23.3% Sodium Chloride (by weight), 76.7% tap water
Group 3: Experimental B	80% Salt Brine from Group 2, 20% Sugar Beet Molasses (Grafschafter Goldsaft)

#### 4. *C. elegans*

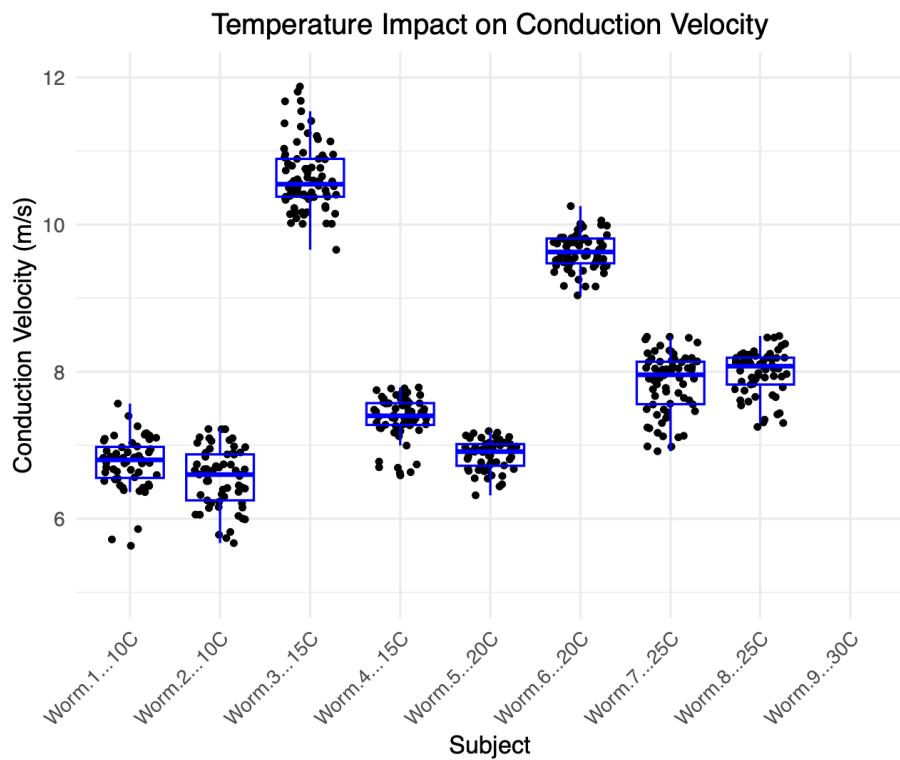
To test whether a species would be repelled by salt brine or beet de-icer, a study of *C. elegans* was performed via chemotaxis behavioral assay, a process by which a petri dish was split into three sections: a start, control (10% ethanol) and experimental. One trial of three different groups was performed with the setup displayed in Figure 4. To confirm the accuracy of the experiment, a paralytic was added to each treatment zone to ensure that no nematode was left in the experimental environment. Benzaldehyde, a known attractant to *C. elegans*, was used as a control to confirm how many nematodes would migrate to a preferred environment.



**Figure 4: Structure of the Chemotaxis Assay**

### III. Data and Analysis

#### a. Temperature

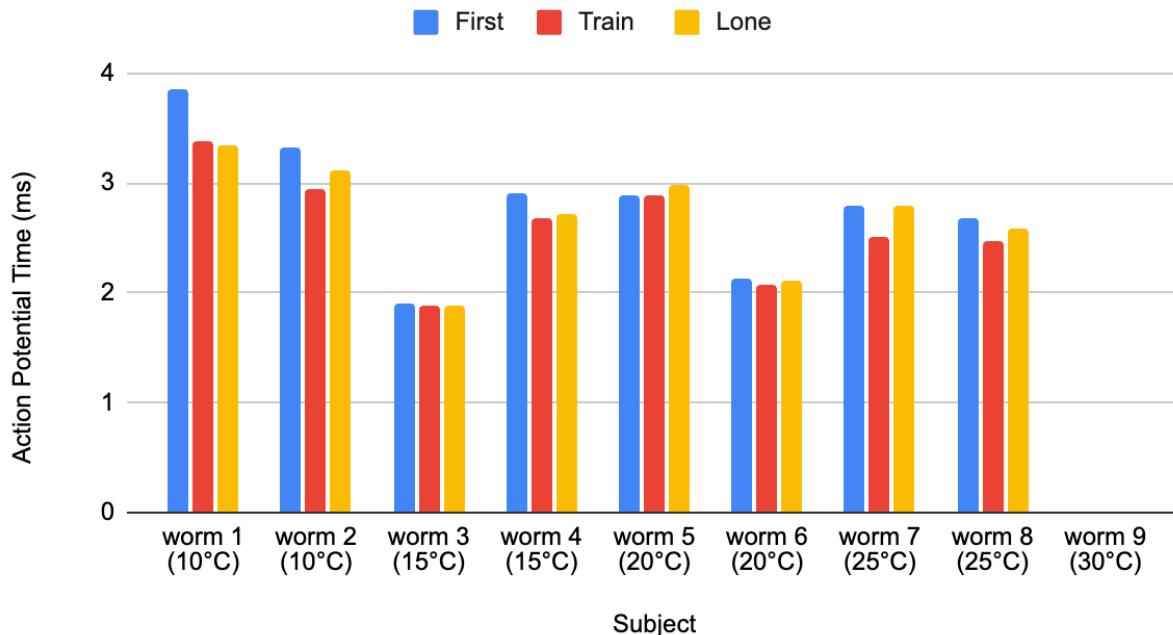


**Figure 5: Conduction velocity of *L. terrestris* in response to chronic temperature exposure**

Chronic exposure to varying temperature groups, increasing in increments of 5°C starting at 10°C, resulted in changes to the conduction velocities of the *L. terrestris* subjects. Figure 5 demonstrates that the subject groups held at lower temperatures tended to have slower average conduction velocities than those held at higher temperatures: the 10°C group containing Worm 1 and Worm 2 had an average group conduction velocity of 6.7727 m/s and the 25°C group containing Worm 7 and Worm 8 had an average group conduction velocity of 8.0375 m/s, as shown in Table 4. This data is in line with our hypothesis that temperature and conduction velocity would have a direct correlation for *L. terrestris*.

In order to produce the most accurate data, a minimum of 55 action potentials were recorded per subject, meaning that for each subject, at least 55 conduction velocities were calculated. From these numerous action potentials, we identified a curious trend where the conduction velocity of the first action potential in a train was consistently slower than the cumulative average conduction velocity calculated for the subject. Conduction velocities calculated from lone action potentials tended to agree with the averages derived from action potentials located in a train. We are uncertain as to why this trend occurred across multiple subjects in different temperature groups but were able to identify that subjects held at lower temperatures seemed to generally have a larger disparity between the action potential times, and therefore conduction velocity, than those held at higher temperatures. A comparison between the time of action potentials at the start of a train, in a train, and alone apart from a train is given in Figure 6 below.

## Average Time of Action Potentials Separated by Cluster



**Figure 6: Average Action Potential times of *L. terrestris* Separated by Locations in Trains**

We observed a qualitative increase in activity among the *L. terrestris* subjects at temperatures closer to their natural environment of 10°C, more so than those maintained at higher temperature environments. The subjects in the temperature groups of 15°C and 10°C all exhibited much higher amounts of activity while submersed in the anesthetic solution than their counterparts in the 20°C and 25°C groups. Similarly, those in lower temperature groups generally required a longer period of time in the anesthetic solution to become fully anesthetized than the subjects kept at higher temperatures. No quantitative data comparing these different responses was recorded, however, meaning that these observations should be regarded as supplemental to the main study on temperature-induced conduction velocity differences.

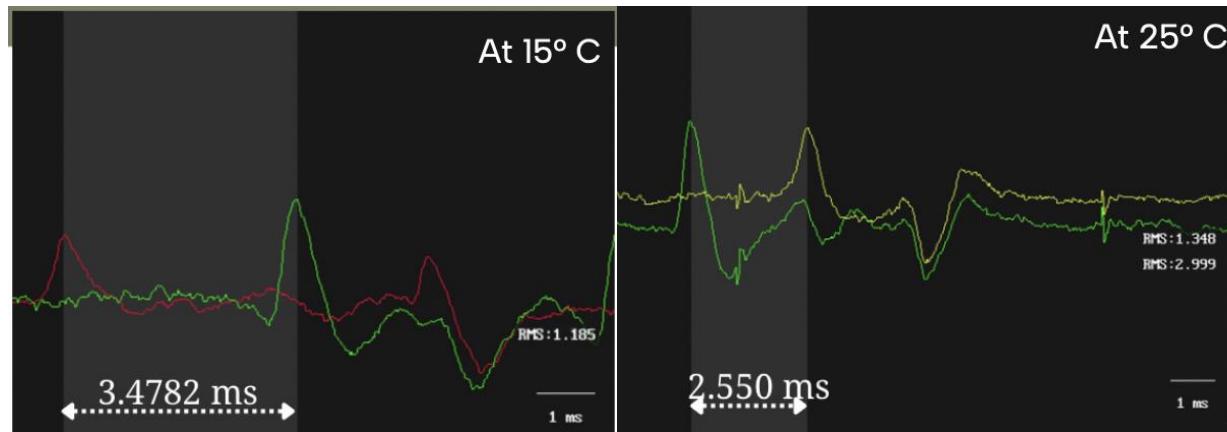
As shown by Figure 5, our data does not fully agree with our hypothesis that conduction velocity would increase with temperature. While the majority of the subjects affirm our hypothesis, worms 3 and 6 at temperatures of 15°C and 20°C respectively were seen to have conduction velocities much higher than expected. Worm 3, part of the 15°C group, had an average conduction velocity of 10.7853 m/s while Worm 4, belonging to the same temperature group, had an average conduction velocity of 7.4852 m/s. Such a large disparity between two worms in the same experimental group was not expected and is indicative that our data is not fully conclusive. This disparity could be explained by the natural variance of conduction velocity between worms due to non-uniform diameters of axons and different degrees of myelin coating the neurons.<sup>31</sup> Specific to Worm 3, the divergence from the expected conduction velocity and the calculated velocity could in part be a cause of the action potentials' irregular shape: no characteristic valleys common to both nodes were observed for any of the action potentials in Worm 3. Another subject that differed from our hypothesis was Worm 9 which was housed at 30°C. From Figure 5, no conduction velocity was measured from Worm 9 due to the temperature in which it was housed proving fatal.

**Table 4: Average individual subject conduction velocities**

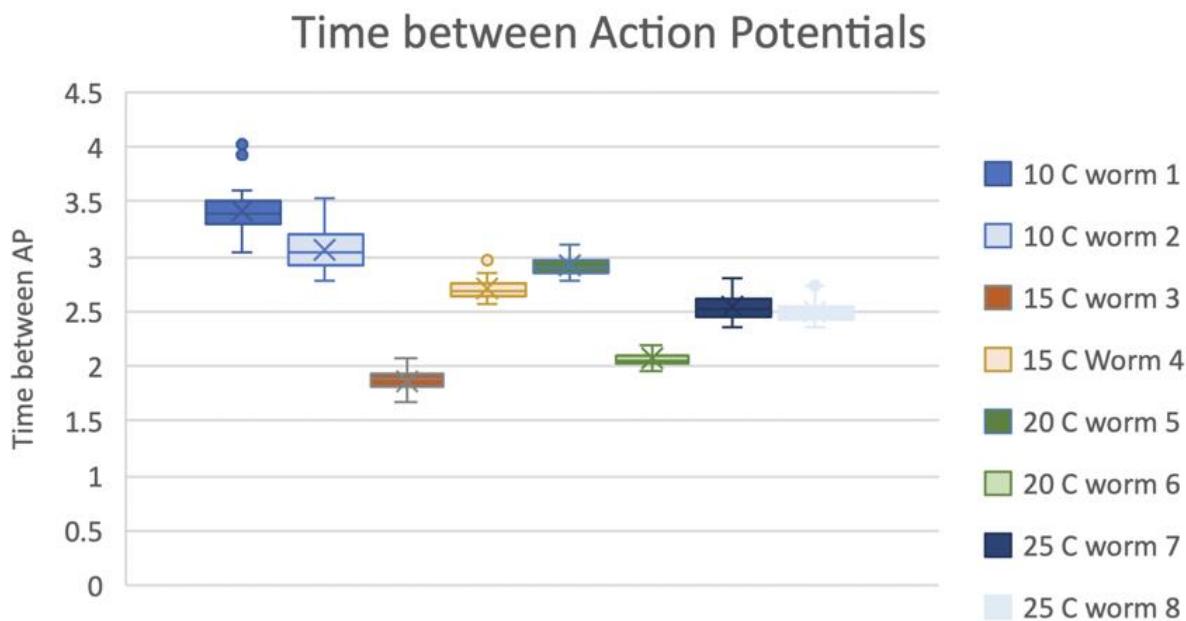
Subject	Average Individual Conduction Velocity (m/s)	Average Group Conduction Velocity (m/s)
Worm 1 (10° C)	6.88	6.77
Worm 2 (10° C)	6.66	
Worm 3 (15° C)	10.78	9.13
Worm 4 (15° C)	7.48	
Worm 5 (20°C)	6.98	8.38
Worm 6 (20° C)	9.78	
Worm 7 (25° C)	7.96	8.04
Worm 8 (25° C)	8.12	
Worm 9 (30° C)	0	0

**Table 5: Statistical Analysis of Chronic Temperature Data from *L. terrestris***

	worm 1 (10°C)	worm 2 (10°C)	worm 3 (15°C)	worm 4 (15°C)	worm 5 (20°C)	worm 6 (20°C)	worm 7 (25°C)	worm 8 (25°C)	worm 9 (30°C)
Median	6.80	6.60	10.55	7.40	6.91	9.63	7.96	8.07	0
Range	1.93	1.55	2.22	1.19	0.87	1.21	1.56	1.24	0
Standard Deviation	0.36	0.40	0.45	0.30	0.20	0.24	0.40	0.31	0

**Figure 7: Time Differences Between Action Potentials in Hot vs Cold Conditions**

The conduction velocity data over incremental increases of temperature suggests that the effects of nerve function can be attributed to the ambient temperature affected by the kinetics of channel gating. As temperature increases, the length between recorded action potentials was shorter when measured in milliseconds. The average distance between peaks in the 15°C group and 25°C group is depicted in both images in Figure 3 which exemplifies clear differences in the time duration of the action potential to travel down the axon as recorded between both nodes. When these action potentials representing the averages of their respective groups are compared in Figure 7, the distance between peaks increased as temperature decreased. It is important to note the distance between both nodes measuring action potential remained constant at 2 cm, so a change in time directly reflects a change in velocity. Our data suggests temperature and conduction velocity may have a direct correlation in *L. terrestris*, and as shown in Figure 8, the time between action potentials decreased as temperature increased. This inverse relationship can be attributed to the sensitivity of action potential firing rates. Action potentials are generated from the transmembrane flux of ions via ion channels. The cell hyperpolarizes temporarily when positive ions can exit the neuron as potassium channels stay open for a little longer. Cooling a neuron, however, causes the ion channels to open more slowly which reduces the speed of the action potential as it travels down the axon further suggesting the reduction in conduction velocity. This makes sense as other studies have found a relationship between warming temperature and an acceleration of the electrogenic sodium pump<sup>32</sup>. Ultimately, our data suggests temperature and conduction velocity may have a direct correlation in *L. terrestris* up to 30° C where data becomes impossible to collect due to the treatment proving fatal.



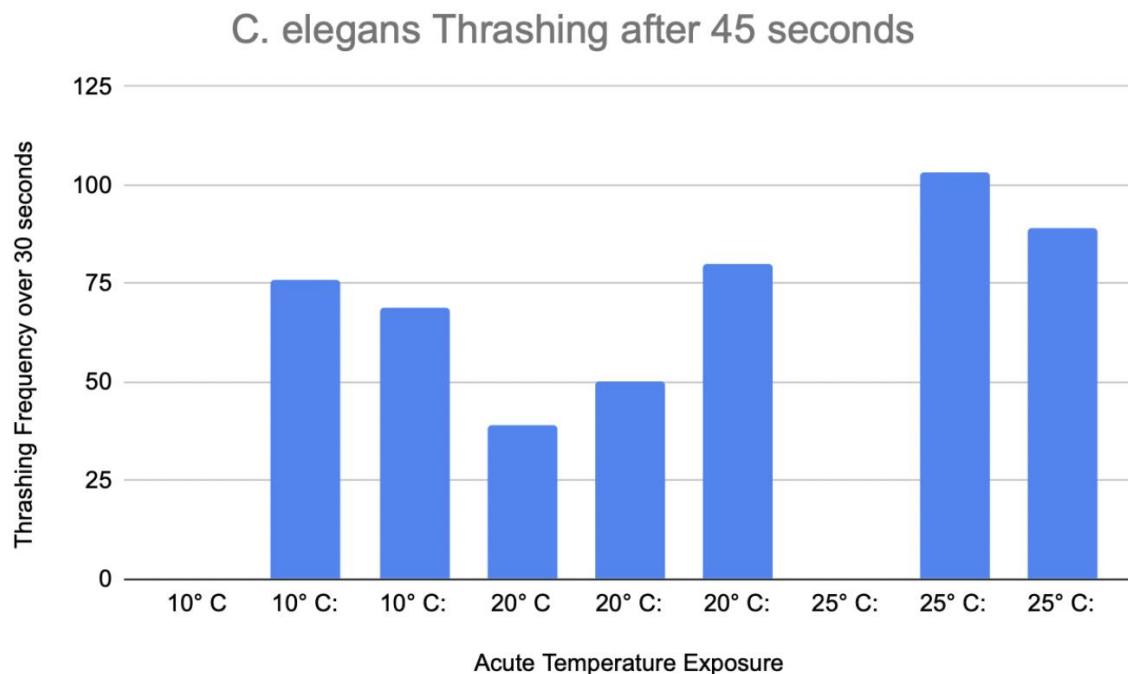
**Figure 8: Average Time Between Action Potentials for *L. terrestris***

In order to further support our hypothesis, *C. elegans* were also tested using acute temperature exposure in a thrashing assay to determine the effect of temperature on neuromuscular activity and neuron function. According to the Nature Reviews Drug Discovery journal, *C. elegans* are often used as model organisms due to the nematode's simplicity and cost-effectiveness.<sup>33</sup> In order to further understand the impact of increasing heat on neuron function, we exposed 3 groups of *C. elegans*, each with three subjects, to 10°C, 20°C, and 25°C respectively, all within the standard laboratory temperature range for *C. elegans*.<sup>34</sup> The *C.*

*elegans* were placed in a 50  $\mu$ l sample of M9 buffer to measure the thrashing frequency within 30 seconds after 45 seconds had passed on the buffer. The buffer was cooled or heated to 10°C and 25°C to induce acute temperature exposure on the *C. elegans* and its effect on neuron function by measuring the mobility of the worms. We hypothesized an increase in mobility as temperature increased due to our findings in the previous experimental assay where we found an increased conduction velocity with warmer chronic temperature exposure.

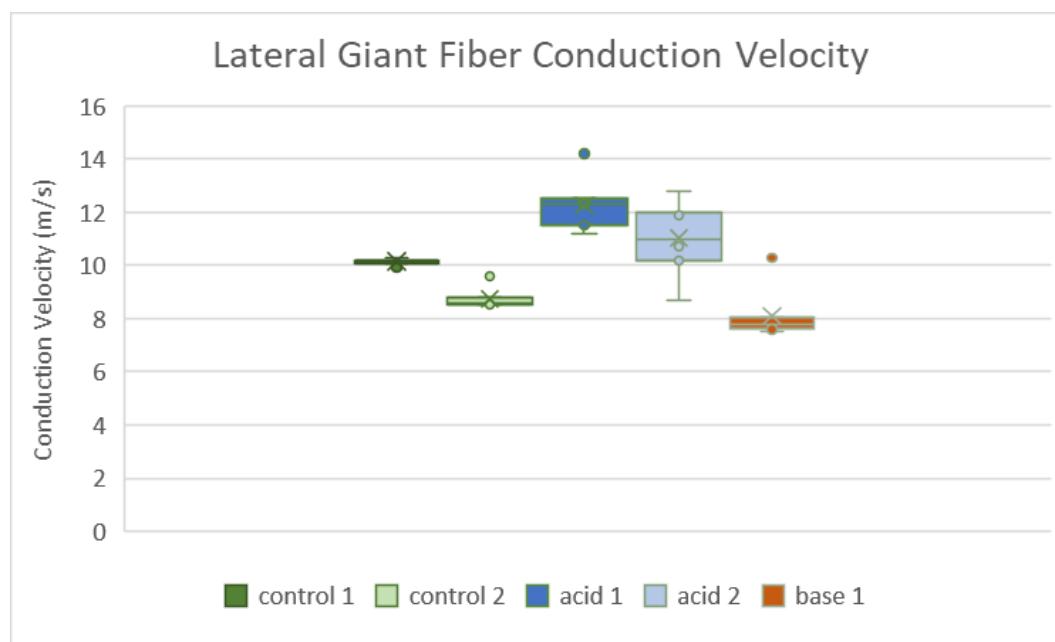
With the exception of three worms, we found data that depicted an increase in thrashing frequency as temperature increased as shown in Figure 9. The first worm in the 10°C group as well as the first worm in the 25°C group did not appear to move within the 30 seconds measured after 45 seconds had passed on the buffer. This result may be due partly to a slower rate of acclimation to the buffer of the specific worm or damage accrued during transfer into the buffer. The individual rate of acclimation may also contribute to the second worm in the 10° C acute temperature exposure test which resulted in a higher thrashing count than the third 10°C worm. Another source of ambiguity may have been that throughout the assay, the 10°C buffer may have heated up closer to the room temperature of 20°C, and in so doing, explaining why the group had similar thrashing frequencies to the 20°C group. Nonetheless, the 25°C group had an average thrashing frequency much higher than any other group, which does support our hypothesis that higher thrashing frequencies correlate with an increase in acute temperature exposure.

**Figure 9: Thrashing Frequency of *C. elegans* in Response to Acute Temperature Exposure**



**b. pH****Table 5: Conduction Velocities of *L. Terrestris* in Control, Acid, and Base Groups**

	Control Worm	Control Worm	Acid Worm 1	Acid Worm 2	Base Worm 1
pH of Soil:	7.5	7.5	4.5	4.5	8.5
Distance between sensors (m):	0.0420	0.0319	0.0642	0.0214	0.0336
Conduction Velocity Trial 1 (m/s):	10.10	7.23	11.19	8.69	10.25
Conduction Velocity Trial 2 (m/s):	10.22	8.06	11.70	10.20	7.54
Conduction Velocity Trial 3 (m/s):	9.92	8.18	11.47	10.71	7.81
Conduction Velocity Trial 4 (m/s):	10.20	7.43	12.28	11.05	7.76
Conduction Velocity Trial 5 (m/s):	10.32	8.59	12.51	11.98	7.68
Conduction Velocity Trial 6 (m/s):	10.23	9.01	12.41	12.77	7.60
Conduction Velocity Trial 7 (m/s):	10.15	8.18	14.20	12.04	8.06
Average Conduction Velocity (m/s):	10.16	8.10	12.25	11.06	8.10

**Figure 10: LGF Conduction Velocities of Acid/Base/Control Worms**

**Table 6: Statistical Analysis of Calculated Conduction Velocities (m/s)**

	Control	Control	Acid Worm 1	Acid Worm 2	Base Worm
Median	10.20	8.18	12.30	11.00	7.76
Range	0.38	1.78	3.00	4.11	2.76
Standard Deviation	0.07	0.45	0.67	1.02	0.76

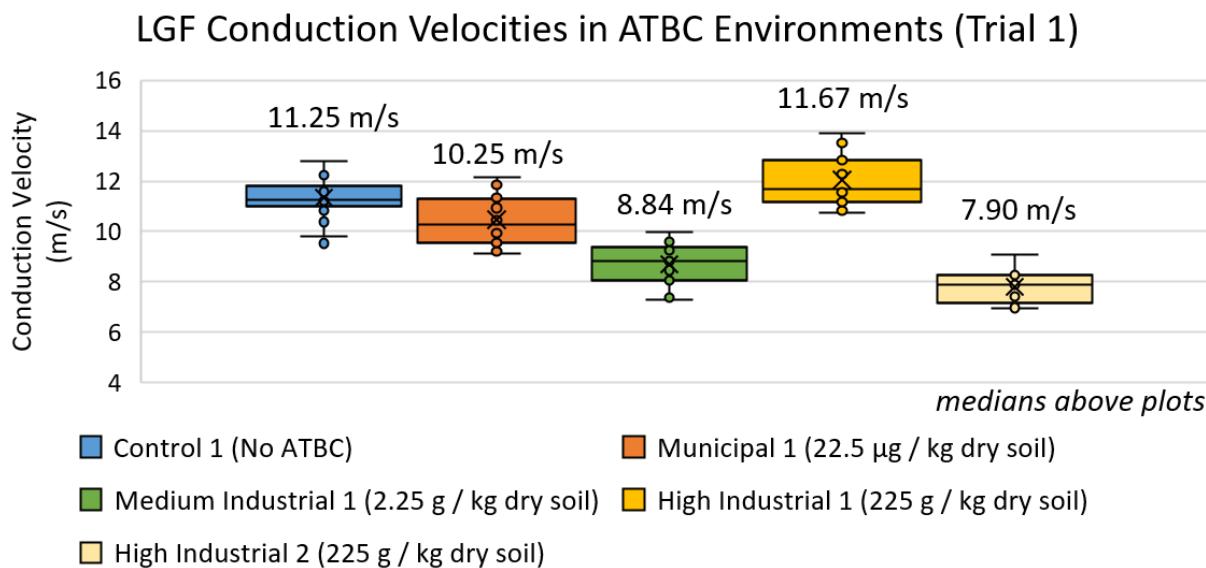
The worm placed in the alkaline environment showed a decrease in conduction velocity compared to the control worms, whereas the worms placed in the acidic environment showed an increase in conduction velocity. The two factors causing this change could either be the distance between the pins or the time between action potentials. As seen in figure 10 and table 5, the distance of the pins is variable to every worm, but even with significantly variable distances, the conduction velocity of the acid worm remained higher than that of the control. In table 5, it shows that one of the worms in the acid environment had a distance of 6.42 cm between pins (electrodes) and another had a distance of 2.14 cm. Despite the difference, both distances still yielded conduction velocities of 12.3 m/s and 11.1 m/s, respectively, which are both greater than the conduction velocities of 10.2 m/s and 8.10 m/s measured from the control worms. This shows that the distance cannot be a reasonable factor for causing the changes in conduction velocity, leaving time as the only factor that is causing the change. Since the acid worm has a higher conduction velocity and distance is not a changing factor, the only way conduction velocity could have increased is if the time between the action potential travelling decreased. An explanation for this change could be due to the increased hydrogen ion concentration outside of the cell due to the acidic environment. This increases the electrochemical gradient, which means there is a more positive charge outside of the neuron. This can cause more sodium ion channels and calcium ion channels to open. More ion channels mean increased flow of positive ions into the cell. This could lead to an increased rate of depolarization. If the depolarization phase of the action potential goes faster, then there can be an overall increase in the rate of action potentials because more potassium channels will open causing repolarization to occur sooner. This supports the fact that the time between the action potential travelling down the nerve cord in the acid worm was decreased which resulted in an increased conduction velocity.

The acid worms showed an increase in conduction velocity but that does not always correlate to more optimal neural function. An acidic environment is harmful, so the increase in conduction velocity could represent hyperactivity and overstimulation in the worms.<sup>35</sup> This goes against our original hypothesis, where we predicted that any variation in pH would lead to a decrease in conduction velocity. For the base treatment, the expected decrease in conduction velocity was seen due to neurodegeneration, which again shows the harmful impact of a change in pH.

One of the biggest problems in this experiment was how the worms seemed to prefer being outside of our treatments. For two different treatment setups, the worms escaped from the soil setups, including the control cups. In the final treatment, however, after properly securing the worms, we were able to get data. One of the basic worms still did not produce visible action potentials which limited the number of basic worms from which we were able to collect data. It is not fully known why the worms kept escaping, but we expect that it was due to an unfavorable environment because of the acid and base or potentially something in the soil.

### c. Phthalate Alternative - ATBC

The earthworms were exposed to different concentrations of ATBC over our experimentation period. Their conduction velocities were measured twice. Below are the box-and-whisker representations of the conduction velocities measured from each worm.

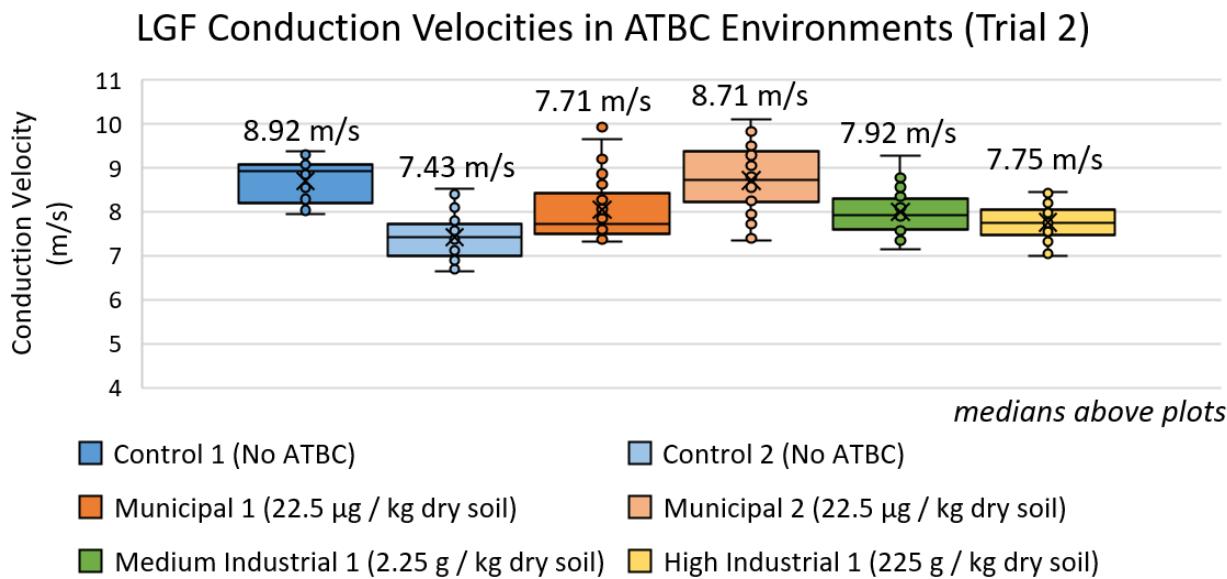


**Figure 11: LGF Conduction Velocities Taken During First Round of Measurements**

**Table 7: Statistical Analysis of First Round Conduction Velocities**

	Control 1	Municipal	Medium	High	High
Median (m/s)	11.25	10.25	8.84	11.67	7.90
Range (m/s)	3.26	3.05	2.70	3.16	2.10
Standard	0.73	0.95	0.79	1.03	0.60

During the first round of measurements, the worms exhibited slower conduction velocities (CVs) when they were exposed to higher concentrations of ATBC. In Figure 11, the control worm had a median CV of 11.25 m/s while the municipal and medium industrial worms had median CVs of 10.25 m/s and 8.84 m/s respectively. In comparison, the high industrial worms had median CVs of 11.67 m/s and 7.90 m/s. Without the one outlier worm treated with a high industrial concentration of ATBC, which exhibited a higher median conduction velocity than the control worm, there was a clear downwards trend. This could be attributed to the worm starting with a higher conduction velocity than any other worm. However, we failed to measure the conduction velocities of all worms before experimentation commenced. The overall trend shown in Figure 11 supports our hypothesis since LGF conduction velocity generally decreased as ATBC concentration increased. This may be due to ATBC's potential to affect glial cell regeneration, reducing glial cells' ability to protect neural cells and leading to neurodegeneration<sup>36</sup>.



**Figure 12: LGF Conduction Velocities Taken During Second Round of Measurements**

**Table 7: Statistical analysis of second round conduction velocities**

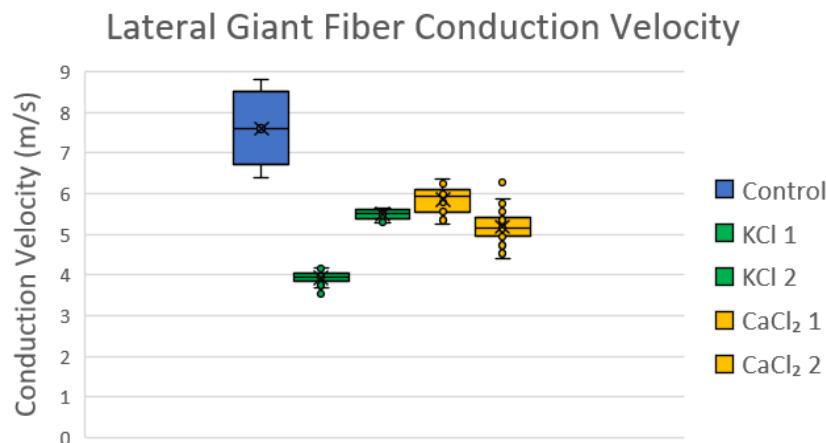
	Control 1	Control 2	Municipal 1	Municipal 2	Medium	High
Median (m/s)	8.92	7.43	7.71	8.71	7.92	7.75
Range (m/s)	1.44	1.87	2.60	2.75	2.13	1.45
Standard	0.45	0.47	0.70	0.74	0.47	0.35

The conduction velocities from the second round of measurements, measured 4 days after the first round, did not yield evident trends as the first round did. Finding trends was nearly impossible because of the variable maintenance conditions of the worms. During previous attempts of the ATBC trials, some worms managed to escape containment due to insufficient precautions. Some worms escaped out of the plastic wrap covers of the paper cups, while others escaped when a paper cup was degraded by a high industrial ATBC subgroup. Therefore, some worms had to be replaced when they were not found in their containers. The worms that did not escape were kept in their respective containers for extended periods of time while the data collection period was restarted. Because the worms that escaped were mainly from a group that was treated with ATBC, the control group would have experienced greater neurodegeneration due to aging and would have lower conduction velocities. There is also variance expected with the worms in the same subgroup since some were exposed longer than others. Two of our worms, one from our medium industrial group and one from our high industrial group, died before action potentials were recorded. Therefore, there was not a sufficient sample size to create a definitive trend. This inhibits our ability to draw conclusions, particularly from the data collected during the second round of measurements.

#### d. Road Salts

**Table 8: Qualitative Analysis of Activity Levels of Worms in Treatments**

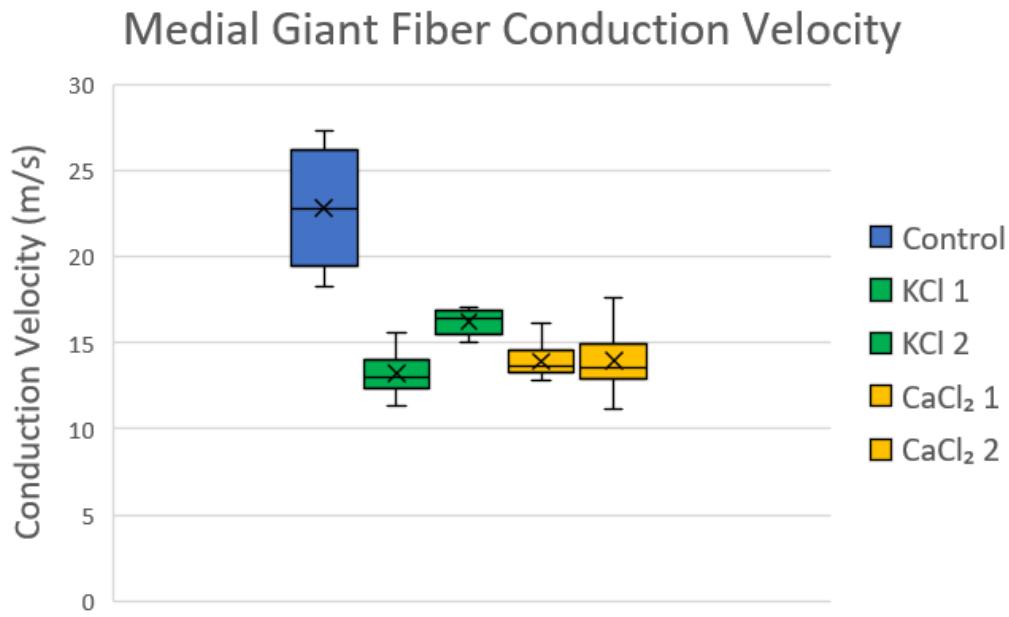
Group of Worm	Activity Level	Post-Treatment
Group 1 (Control)	Baseline	Alive and healthy
Group 2 (1% KCl)	Baseline	Alive and sluggish
Group 3 (1% CaCl <sub>2</sub> )	Baseline	Alive and sluggish
Group 4 (3.5% & 26% KCl)	Moderate-Heightened Levels of Activity	Dead and dry
Group 5 (3.5% & 26% CaCl <sub>2</sub> )	Heightened Levels of Activity	Dead, dry, & hemorrhaged



**Figure 13: LGF Conduction Velocity**

**Table 9: Statistical Analysis of LGF Conduction Velocities**

	Control	KCl 1	KCl 2	CaCl <sub>2</sub> 1	CaCl <sub>2</sub> 2
Median (m/s)	7.60	3.95	5.52	5.94	5.15
Range (m/s)	2.40	0.65	0.35	1.10	1.90
Standard	1.20	0.17	0.12	0.34	0.41
Mean (m/s)	7.60	3.74	5.48	5.86	5.18
Mean Group (m/s)	7.60	4.71		5.52	

**Figure 14: MGF Conduction Velocity****Table 10: Statistical Analysis of MGF Conduction Velocities**

	Control	KCl 1	KCl 2	CaCl <sub>2</sub> 1	CaCl <sub>2</sub> 2
Median (m/s)	22.80	12.96	16.45	13.62	13.59
Range (m/s)	9.00	4.31	1.96	3.39	6.49
Standard	4.50	1.15	0.69	0.83	1.55
Mean (m/s)	22.80	13.25	16.23	13.92	13.42
Mean Group (m/s)	22.80	14.74		13.67	

After being exposed to the treatments for four days, the conduction velocities of the worms were recorded. In the first two trials with concentrations of 26% and 3.5% concentrations, the KCl and CaCl<sub>2</sub> worms were dead, visibly drier, and thinner than the worms in the control group. Worms exposed to 1% concentrations looked healthier than in the first and second trials but were still drier and thinner than the control worms. In the first two trials, the worms exposed to the salts were much more active immediately after being exposed to the salts than the worms in the control group. As the worms in the first two trials were dead, it was not possible to record conduction velocities for those worms. As predicted in our hypothesis, the worms exposed to 1% concentrations of KCl and CaCl<sub>2</sub> had slower conduction velocities. As shown in Table 9, the four worms exposed to the road salts demonstrated slower LGF conduction velocities than the control as the KCl group had an average conduction velocity of 4.71 m/s, the CaCl<sub>2</sub> group had an average conduction velocity of 5.52 m/s, and the control group had an average conduction velocity of 7.6 m/s. Table 10 shows that the same is true for the MGF conduction velocities with average conduction velocities of 14.74 m/s, 13.67 m/s, and 22.8 m/s for the KCl group, CaCl<sub>2</sub> group, and the control, respectively. Therefore, road salts slowed conduction velocities in earthworms which is a result of neural degradation.

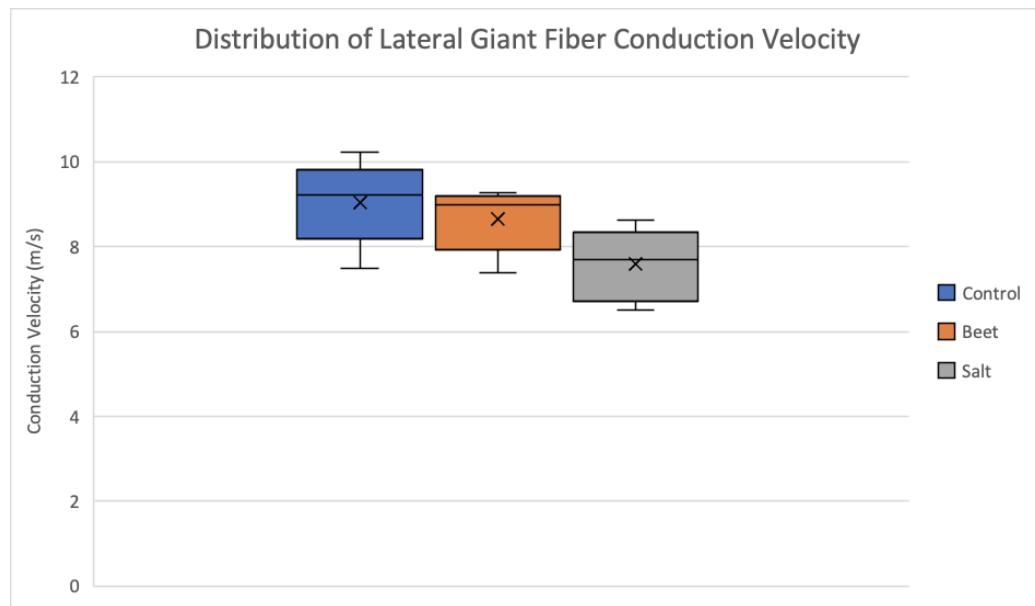
Some sources of error include that individual earthworms have individual differences that cause variations between their conduction velocities which can make comparisons between different worms harder to interpret with a smaller sample size. Other sources of error could be from worms taking varying amounts of time to be fully anesthetized which is very important to measuring accurate action potentials. The worms also were pinned and unpinned multiple times causing trauma, which could impact their neurological function. There were also only two worms per experimental group meaning that if one worm was an outlier and had poor neurological function the data would be heavily influenced.

### e. Sugar Beet Additives

The three separate groups each consisting of three worms were exposed to either no treatment, the salt brine de-icer, or the sugar beet de-icer. This experiment was completed twice, using a total of eighteen worms. Figure 15 and Table 11 display the data for this experiment.

A clear trend can be viewed, with the more salt the solution contains, the lower the conduction velocity present in *L.terrestris*, which is supported by our data. For both the mean and median conduction velocities, worms treated with the salt brine de-icer solution demonstrated a much bigger decrease from the control worm conduction velocity (mean:  $\Delta CD = -1.4533$ , median:  $\Delta CD = -0.979$ ) as compared to the decrease in worms treated with sugar beet de-icer (mean:  $\Delta CD = -0.4032$ , median:  $\Delta CD = -0.106$ ). In addition, the sugar beet de-icer is statistically still in the range of the control, implying that the sugar beet juice de-icer is more effective at limiting environmental damage than salt brine alone. It is suggested that future research be completed to determine if the sugar beet de-icer be tested in other conditions to fully determine if the de-icer is truly more environmentally friendly than salt brine.

Sources of error may have included trauma to the worms, incomplete or overdone anesthetization, and inconsistencies in the simulated precipitation.



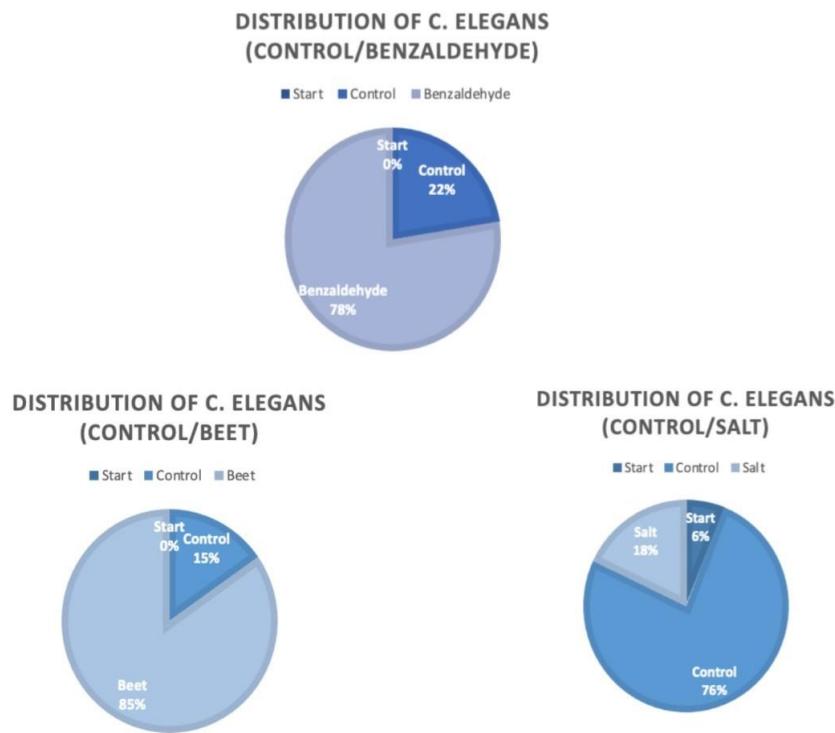
**Figure 15: Distribution of Lateral Giant Fiber Conduction Velocity**

**Table 11: Statistical Analysis of Beet/Salt De-Icer Conduction Velocity**

	Control	Beet De-Icer	Salt De-Icer
n	5	5	6
Mean	9.05	8.64	7.59
SD	1.00	0.77	0.82
Median	9.21	9.00	7.69

We used a chemotaxis assay to test whether *C. elegans* would be repelled from any form of salt brine. A control of 10% ethanol was used on all plates, in addition to a dedicated control plate consisting of benzaldehyde, a known attractant, and ethanol. It was determined that the nematodes greatly preferred the beet de-icer, even more so than the benzaldehyde, as shown by the data in Figure 16.

Sources of error include that only one trial was conducted and only one person counted the nematodes, meaning that a miscount could have occurred.

**Figure 16: Distribution of C. elegans in the Chemotaxis Assay**

## IV. Conclusion

### a. Error Analysis

Many subgroups experienced variation in the conduction velocities of different worms in the same experimental groups. This is because earthworms have inherent differences in their conduction velocities before any treatment is administered. When minute changes in conduction velocities are being measured, these inherent differences could completely invalidate findings. To combat these differences, we could measure the initial velocities of the worms in each subgroup and compare the change in average velocity after a set number of days. An additional adjustment could have been to use a much larger data pool. By increasing the number of worms to around 60 or more<sup>11</sup>, the variation in conduction velocities and anesthetization times would average out to become far more insignificant. This would be a much more accurate measure of neurodegeneration than comparison of arbitrary worms placed in given conditions.

Because the purpose of this study was to determine the effect of different environmental factors on neurodegeneration, being able to compare the treatments of different subgroups would have been convenient and interesting. However, many experimental factors varied between subgroups. For example, different subgroups bathed earthworms in the 10% EtOH for different times to fully anesthetize the worms. Different exposures could lead to sluggish conduction velocities, making results incomparable. Also, the setups of the earthworm environment were different. For example, the subgroup that tested chloride-based salts stored their worms in paper cups while the group that tested salt and beet mixtures stored their worms in plastic tubs and exposed worms via an asphalt runoff system. These differences prevented us from comparing the results of these groups to find optimal road salting configurations. Overall, more standardization would improve the experiment.

### b. Future Applications

These findings can be used to further explore the ramifications that humans have on their environment, particularly for factors such as plasticizers and sugar beet additives that have not been researched thoroughly. The relevance of the experiments can be determined in humans as worms serve as model organisms for the neurological function of humans. Likewise, the worms can serve as an indicator of the environment's health since they are a foundational part of it.

Upon exposure to the five environmental factors studied, we observed a dramatic impact on conduction velocity for *L. terrestris*, as well as a comparable change in behavior for *C. elegans*. This data suggests that, as these pollutants become more abundant, the global ecosystem will be similarly negatively impacted. Given the worsening state of the global ecosystem, studies such as this will be increasingly vital in identifying which human-related pollutants have adverse effects on organisms' health. Temperatures are expected to continue increasing, pH levels will continue to change, and microplastics are likely to penetrate further into the environment. Consequently, organisms from every ecosystem will increasingly face the repercussions of unchecked industrial activity. Identifying the harm posed by our unsustainable practices is the first step toward mitigating, and eventually ameliorating, these behaviors.

### c. Acknowledgements

We would like to acknowledge and deeply thank Dr. Emily Drill and TA Michael Sobol as well as Carnegie Mellon University, the Department of Biology, Dr. Barry Luokkala, Ms. Melissa Lessure, Dr. Carrie Doonan,

the PGSS Alumni Campaign, and parents and guardians for providing the resources and support that gave us this opportunity for research. This would not have been possible without them.

## V. References

1. Steven Cohen, "The growing awareness and prominence of Environmental Sustainability," 19 September, 2022, <https://news.climate.columbia.edu/2022/09/19/the-growing-awareness-and-prominence-of-environmental-sustainability/>
2. Rima Habre, David C. Dorman, Jonathan Abbott, William P. Bahnfleth, Ellison Carter, Delphine Farmer, Gillian Gawne-Mittelstaedt, Allen H. Goldstein, Vicki H. Grassian, Glenn Morrison, Jordan Peccia, Dustin Poppendieck, Kimberly A. Prather, Manabu Shiraiwa, Heather M. Stapleton, Meredith Williams, and Megan E. Harries, "Why Indoor Chemistry Matters: A National Academies Consensus Report," *Environmental Science & Technology* **Volume 56**, (Issue 15), n.p, (14 July 2022).
3. C. Cvitanovic, J. McDonald, A.J. Hobday, "From science to action: principles for undertaking environmental research that enables knowledge exchange and evidence-based decision-making," *Journal of Environmental Management* **Volume 183**, (Issue 3), 864-874, (1 December 2016)
4. Kiernan, Matthew, Cikurel, Katia, Bostockm, Hugh, "Effects of temperature on the excitability properties of human motor axons," *Oxford Academic* Volume 124, April 2001, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6944501/>
5. Habibi L, Perry G, Mahmoudi M. Global warming and neurodegenerative disorder: speculations on their linkage," *Bioimpacts*, 167-170, (2014)
6. Subaraja M, Vanisree AJ. Cerebral ganglion variations and movement behaviors of the Lumbricus terrestris on exposure to neurotoxin. *Ann Neurosci*, (2015)
7. Experiment: Comparing Speeds of Two Nerve Fiber Sizes, <https://backyardbrains.com/experiments/comparingnervespeed>
8. Ericsson AC, Crim MJ, Franklin CL, "A brief history of animal modeling," *Mo Med* **Volume 110**, (3), 201-205, (May-June 2013).
9. Steven M. Chrysafides, Stephen J. Bordes, Sandeep Sharma, Physiology, Resting Potential, StatPearls Publishing; 2023 Jan-, Florida
10. Michael H. Grider, Rishita Jessu, Rian Kabir, Physiology, Resting Potential ,StatPearls Publishing, 2023 Jan-, Florida
11. Shannon, K. M., Gage, G. J., & Jankovic, A., "Portable conduction velocity experiments using earthworms for the college and high school neuroscience teaching laboratory," *American Physiological Society* **Volume 33**, (Issue 1), 62, (March 2014)
12. Dahlman, R. L. A. L. (n.d.). *Climate change: Global temperature*. NOAA Climate.gov. [https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=Earth's%20temperature%20has%20risen%20by,0.18%C2%B0%20C\)%20per%20decade.](https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=Earth's%20temperature%20has%20risen%20by,0.18%C2%B0%20C)%20per%20decade.)
13. Lee, S.-J., & Kenyon, C. (2009, May 12). *Regulation of the longevity response to temperature by thermosensory neurons in Caenorhabditis elegans*. Current biology : CB. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2868911/>
14. SA;, K. M. M. P. (n.d.). *Physiological responses to temperature and haeme synthesis modifiers in Earthworm Lumbricus terrestris (Annelida: Oligochaeta)*. Environmental toxicology. <https://pubmed.ncbi.nlm.nih.gov/20725936/>
15. Glenn Suter II, Susan Cormier, Kate Schofield, Jaime Gilliam, Chad Barbour, "Sources, Stressors & Responses," *CADDIS*, **Volume 2**, March 2023

16. Brinkman, J. E., & Sharma, S., *Physiology, Metabolic Alkalosis*, (StatPearls Publishing, Treasure Island, 2023)
17. J. Hou, Y.Qian, G. Liu and R. Dong, "The influence of temperature, pH and C/N ratio on the growth and survival of earthworms in municipal solid waste" *Agricultural Engineering International: the CIGR Ejournal*, November, 2005
18. Hlisníková H, Petrovičová I, Kolena B, Šidlovská M, Sirotkin A, "Effects and mechanisms of phthalates' action on reproductive processes and reproductive health: a literature review", *International Journal of Environmental Research and Public Health Volume 17*, (Issue 18), (September 18, 2020)
19. Shisan Xu, Huan Zhang, Ping-Chieh Pao, Audrey Lee, Jun Wang, Yu Suen Chan, Francis A.M. Manno III, Shun Wan Chan, Shuk Han Cheng, Xueping Chen, "Exposure to phthalates impaired neurodevelopment through estrogenic effects and induced DNA damage in neurons," *Aquatic Toxicology Volume 222*, (0166-445X), March 10, 2020
20. American Society for Biochemistry and Molecular Biology, Phthalate alternative may harm brain development and health, March 28, 2023,  
<https://www.sciencedaily.com/releases/2023/03/230328145501.htm>
21. Pennsylvania Department of Transportation, "Qualified Products List (QPL) For Winter Road Maintenance," Pennsylvania, 2021,  
[https://gis.penndot.gov/BPR\\_PDF\\_FILES/Documents/LTAP/TechSheets/TS\\_209\\_Winter\\_Road\\_Maintenance.pdf](https://gis.penndot.gov/BPR_PDF_FILES/Documents/LTAP/TechSheets/TS_209_Winter_Road_Maintenance.pdf)
22. Victoria R. Kelly, Stuart E. G. Findlay, William H. Schlesinger, Kirsten Menking, Allison Morrill Chatrchyan, "Road salt moving toward the solution." *The Cary Institute of Ecosystem Studies* (December 2010)
23. Shenandoah Briere, Officials boast salt science, plow blades as cost, environmental savers, Virginia, February 2023
24. Patricia Leigh Gillis, Joseph Salerno, C. James Bennett, Yaryna Kudla, and Margot Smith, "The Relative Toxicity of Road Salt Alternatives to Freshwater Mussels; Examining the Potential Risk of Eco-Friendly De-icing Products to Sensitive Aquatic Species," *ACS EST Water*, 1, (7), 1628–1636, (2021).
25. The Missouri Department of Transportation, "Sugar Beets", Missouri, ND,  
<https://www.modot.org/sugar-beets#:~:text=We%20got%20the%20beet!,icy%20or%20snow%20packed%20surfaces.>
26. Miranda Meehan, Thomas DeSutter, Kevin Sedivec, Chris Augustin, Annalie Peterson, "Environmental impacts of brine (produced water)," (March 2023)
27. Soil health - pH, USDA,  
[https://www.nrcs.usda.gov/sites/default/files/2022-11/pH%20-%20Soil%20Health%20Guide\\_0.pdf](https://www.nrcs.usda.gov/sites/default/files/2022-11/pH%20-%20Soil%20Health%20Guide_0.pdf)
28. Bloom City, pH Up,  
[https://cdn.shopify.com/s/files/1/1576/2437/files/pH\\_Up\\_Label.pdf?v=1625417054](https://cdn.shopify.com/s/files/1/1576/2437/files/pH_Up_Label.pdf?v=1625417054); Bloom City, pH Down, [https://cdn.shopify.com/s/files/1/1576/2437/files/pH\\_Down\\_Label.pdf?v=1625417054](https://cdn.shopify.com/s/files/1/1576/2437/files/pH_Down_Label.pdf?v=1625417054)
29. B.I. Chaudhary, C.L. Liotta, J.M. Cogen, M. Gilbert, "Plasticized PVC", Materials Science and Materials Engineering, page 7114–7117, 2001
30. Frederick Büks, Martin Kaupenjohann, "Global concentrations of microplastics in soils - a review." *Soil Volume 5*, (issue 2), 2020
31. De Col, R., Messlinger, K., & Carr, R. W. (2008, February 15). *Conduction velocity is regulated by sodium channel inactivation in unmyelinated axons innervating the rat cranial meninges*. The Journal of physiology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2375633/>
32. Matthew C. Kiernan, Katia Cikurel, Hugh Bostock, "Effects of temperature on excitability properties of human motor axons," *Brain Volume 124*, (Issue 4), (April 2001)

33. Titus Kaletta, Michael O. Hengartner, "Finding function in novel targets: *C. elegans* as a model organism," *Nature Reviews Drug Discovery*, (May 2006)
34. Gómez-Orte, E., Cornes, E., Zheleva, A., Sáenz-Narciso, B., de Toro, M., Iñiguez, M., López, R., San-Juan, J. F., Ezcurra, B., Sacristán, B., Sánchez-Blanco, A., Cerón, J., & Cabello, J. "Effect of the diet type and temperature on the *C. elegans* transcriptome," *Oncotarget Volume 9*, (Issue 11), (December 2017)
35. Sokoll, M. D., & Thesleff, S., "Effects of pH and uranyl ions on action potential generation and acetylcholine sensitivity of skeletal muscle," *European journal of pharmacology*, **Volume 4**, (Issue 1), 71–76, (1968)
36. Public Affairs, CWU biological sciences graduate student earns praise for research findings, Washington, April 13, 2023, <https://www.cwu.edu/news/campus-and-community/cwu-biological-sciences-graduate-student-earns-praise-research-findings#:~:text=The%20findings%20suggest%20that%20ATBC,to%20neurodegeneration%20and%20accelerated%20aging>

**BIOLOGICAL  
PHYSICS  
TEAM PROJECT**



# Determining the Secondary Structure of a Cyclic Antimicrobial Peptide Using Circular Dichroism

Carter Rogers, Arjun Suryawanshi, Shivani Umesh, Hari Viswanathan, Megan Xie, and Valerie Xu

## Abstract

Antibiotics are losing efficiency in the face of the rapid emergence of drug-resistant bacteria. The development of antimicrobial peptides (AMPs) provides a novel approach that may remedy the antibiotic resistance crisis. The cyclic AMP, CE-0005B, composed of arginine, tryptophan, and valine, was analyzed using circular dichroism (CD) to elucidate secondary structures in Gram-negative bacterial lipid model membranes. Using the Levenberg-Marquardt algorithm, the ellipticity curve from CD was fit according to four motifs:  $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn, and random coil. Minimum inhibitory concentrations (MICs), antimicrobial activity, and eukaryotic cell toxicity were examined relative to the secondary structure compositions. Analysis of the fitted CE-0005B curves indicates a high percentage of  $\beta$ -sheet and random coil structures. MIC and toxicity data suggest that CE-0005B has high efficacy in killing bacteria and low hemolytic activity, making it a promising candidate for further research and potential clinical application.

## I. Introduction

### A. Antibiotic Resistance and Novel Methods of Treatment

There is an increasing emergence of antimicrobial-resistant pathogens. Particularly, antibiotic-resistant bacteria have contributed to a decline in the efficacy of established antibiotic treatments.<sup>1</sup> Since their first modern implementation in the early 20th century, antibiotics have served as a principal form of treatment for bacterial infections.<sup>2</sup> Today, it is estimated that ~210 million antibiotic prescriptions are filled yearly in the United States. However, due to factors such as diagnostic uncertainty, 1 in 7 of these prescriptions are believed to be unnecessary and potentially harmful to patients.<sup>3,4</sup> The overprescription and poor stewardship of antibiotics have both contributed to the rapid development of antimicrobial resistance (AMR). AMR occurs when bacteria undergo genetic changes that render them resistant to specific antibiotic treatments.<sup>5</sup> Treating bacterial populations with antibiotics selects for bacteria that have developed defense mechanisms against antibiotics. Over time, the antibiotic-resistant bacteria proliferate, resulting in a population of largely resistant bacteria.<sup>6</sup>

Each year in the United States, it is estimated that there are more than 2 million infections and 35,000 deaths related to multidrug-resistant (MDR) bacterial infections, resulting in an annual cost of ~20 billion dollars for the US health system.<sup>7,8</sup> Both economic and healthcare-related factors highlight the importance of novel alternatives for antibiotics that function effectively without inducing AMR or adverse effects in patients. Considering these factors, antimicrobial peptides present themselves as a promising candidate to address antibiotic resistance. AMPs are a class of small peptides that exist naturally and function in the innate immune system of organisms. AMPs have existed in nature for millions of years without evoking bacterial resistance. The low development of bacterial resistance associated with AMPs can be attributed to their targeting the cell membrane rather than metabolic pathways. Additionally, AMPs are considered to be less toxic as they are broken down into amino acids, while conventional antibiotics can generate harmful

metabolites.<sup>9,10</sup> The slower development of antibiotic resistance and limited toxicity shown by researchers present AMPs as a viable alternative to antibiotic treatments.

## B. Protein Secondary Structures

Proteins begin as primary structures, which are sequences of amino acids linked via peptide bonds. However, various folds occur between different regions of amino acids, resulting in complex secondary, tertiary, and quaternary structures. In secondary structures, hydrogen bonds form between the amino and carboxyl groups of different amino acids along the protein backbone. Relevant to this study are the four main types of secondary structures:  $\alpha$ -helices,  $\beta$ -sheets,  $\beta$ -turns, and random coils.

The  $\alpha$ -helix secondary structure is a right-handed helical coil held together by hydrogen bonds between every fourth amino acid. Transmembrane proteins contain  $\alpha$ -helices exhibiting hydrophobic properties. This allows transmembrane proteins to remain stable in the cell membrane.<sup>11</sup> The  $\beta$ -sheet secondary structure consists of two regions of a polypeptide chain that lie side by side and are bound together by hydrogen bonds. Structural turns occur when a protein chain requires a change of direction to connect other elements of a secondary structure. Specifically,  $\beta$ -turns resulting from a change of direction occur in the space of four residues.<sup>12</sup> While random coil is not a true secondary structure, it does describe a class of structures that do not conform to the three previous classes of secondary structures mentioned above.

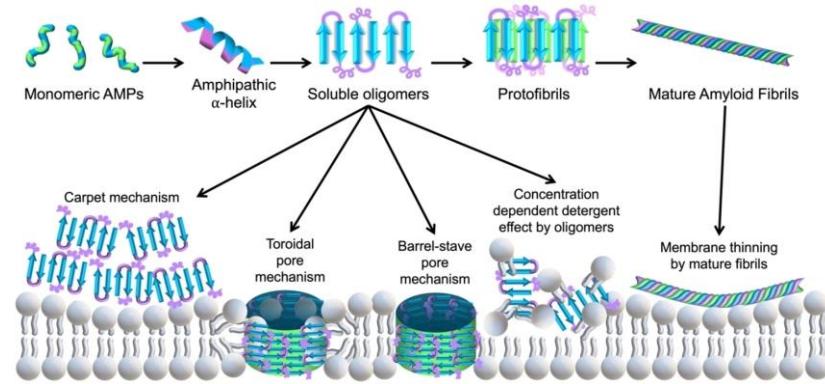
## C. AMP Structures and Mechanisms of Action

In recent years, antimicrobial peptides (AMPs) have shown promising results in addressing antibiotic resistance.<sup>13,14,15</sup> They target highly vital and conserved portions of the bacterial membrane rather than bacterial metabolic pathways, thus reducing the effects of bacterial evolution and natural selection.<sup>18</sup> To ensure the safety of AMP use as an alternative to antibiotics, researchers are currently attempting to elucidate their structures and mechanisms of action in order to translate their findings into clinical applications.

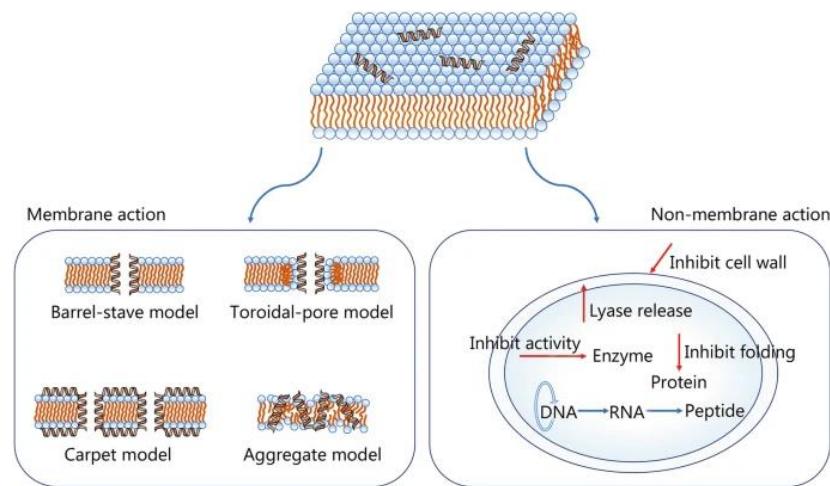
The structure of AMPs often changes after interacting with lipid model membranes (LMMs). The purpose of LMMs is to model interactions between AMPs and bacterial or human eukaryotic cells and ensure that the AMPs have high efficacy and low mammalian cell toxicity. Since AMPs display amphipathicity, their hydrophilic and hydrophobic moieties interact with the hydrophilic heads and hydrophobic tails of phospholipids in the cell membrane. During this interaction, these peptides tend to undergo a conformational transition, folding into secondary structures that disrupt lipid membranes (Fig. 1).<sup>16</sup> Their cationic properties also allow for targeting specificity that prevents autoimmune responses. While most mammalian membranes consist of neutrally-charged phosphatidylcholine (PC) and phosphatidylethanolamine (PE), bacteria have anionic phospholipids, including phosphatidylglycerol (POPG) and cardiolipin (CL). Since the majority of AMPs contain positively charged amino acids such as arginine and lysine (Fig. 2), electrostatic attractions occur between the AMPs and bacterial phospholipids, allowing for specific targeting to bacterial cells only. These attractions allow AMPs to accumulate on the surface of bacterial membranes to ultimately induce bacterial cell death.

Gram-positive and Gram-negative bacteria differ in their membrane compositions. While Gram-positive bacteria contain a thick peptidoglycan layer and a cytoplasmic membrane, Gram-negative bacteria have two phospholipid membranes separated by a periplasmic space with a thin peptidoglycan layer. Previous studies have established that secondary structures of AMPs are dependent on the type of bacteria and membrane composition. Thus, the interactions of AMPs with Gram-negative, Gram-positive, and eukaryotic cells differ.

While the exact mechanisms of action are unclear, researchers have defined four main models to predict AMP action: the barrel-stave model, toroidal-pore model, carpet model, and aggregate model (Fig. 2).<sup>15,17</sup>



**Figure 1: Structural transformations of AMPs upon membrane contact.**



**Figure 2: Membrane and intracellular actions of AMPs on bacterial cells.**

In the barrel-stave model, the peptides interact laterally with the phospholipids. The hydrophobic portions of the AMP interact with the nonpolar phospholipid tails while hydrophilic moieties of the AMPs face inwards, creating a transmembrane pore. The toroidal-pore model is similar to the barrel-stave model, but instead of creating a hole in the membrane, the AMP induces lipid membrane deformation, pulling the phospholipid heads inwards to create a toroidal pore complex. Perforation of the bacterial membrane results in the exchange of ions and water, ultimately leading to a loss of transmembrane potential and cell death. AMPs can also cause complete membrane lysis through the carpet model, in which high concentrations of AMPs insert into the membrane and disrupt the membrane arrangement, leading to micellization. Micelle formation also occurs in the aggregate model, but the aggregate model allows AMPs to enter the cytoplasm, resulting in intracellular inhibition of nucleic acid and protein synthesis, enzymatic activity, and bacterial cell wall synthesis (Fig. 2).

#### D. Evaluation of Existing AMPs

Recent developments have shown that cationic AMPs can increase their selectivity in bacterial cell targeting through specific combinations of amino acids. For example, the chimeric peptide assembles into nanofibers

at pH 7.4 and is transformable in the acidic biofilm-infected microenvironment at pH 5.0.<sup>19</sup> This was achieved by placing positively charged arginine (Arg, R) amino acids on the helix polar face and the hydrophobic valine (Val, V) amino acids on the helix nonpolar face. This composition of amino acids resulted in a size reduction as well as an increased charge. These changes improved the penetration into bacterial biofilms in a mechanism dominated by membrane perforation.

For example, a study developed peptide derivative WLBU2 composed of R (13 residues) and V (8 residues), with 3 Tryptophan (W) residues in the hydrophobic face separated by at least 7 amino acids. This specific structure was engineered to form an ideal helical and amphipathic structure to increase the likelihood of optimal charge and hydrophobic densities.<sup>20</sup> In this study, WLBU2 was better able to eliminate the bacterial biofilm without toxicity to the host cells, and the AMP displayed potent efficacy against the bacterial biofilm.

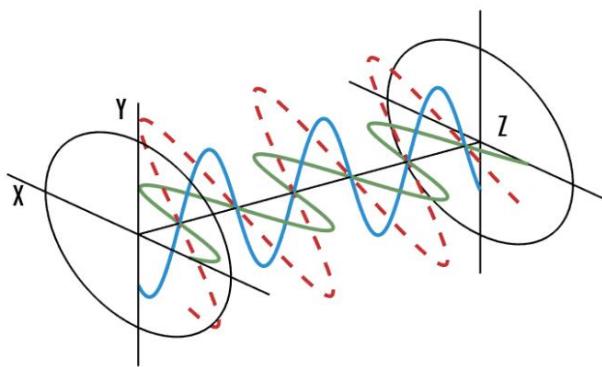
Introducing W on the hydrophobic side of the peptide may have effects on the AMP's toxicity to host cells and efficacy in eliminating the bacteria. W has been known to be prevalent in "penetratins", or regions of peptides that can likely cross membranes. Deslouches et al. have shown that including the W residue both enhances the potency of the AMP against a range of Gram-positive and Gram-negative bacteria and decreases sensitivity to salt and to serum.<sup>20,21</sup>

An increasing challenge in AMP design is to increase antibacterial potency without the risk of toxicity to the host. The design assumed a perfect helix with all positively charged residues on one face while all the hydrophobic residues were on the opposite face.<sup>20</sup> While using W solely in the hydrophobic domain, it not only enhances antimicrobial functions but also reduces host toxicity. This is due to its high hydrophobicity and bulky indole ring. Mitigating the toxicity to the host cells required V to be added to the hydrophobic domain.<sup>24</sup> This is due to its lower hydrophobicity compared to W.

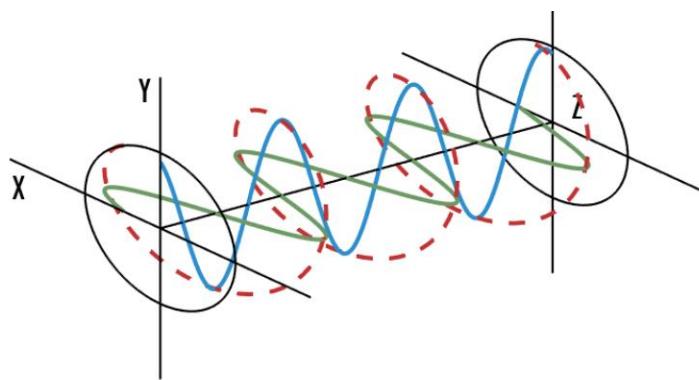
E2-35, an AMP used in this study, was obtained from Tristram-Nagle et al. When looking at minimum inhibitory concentrations (MIC) of AMPs, the higher  $\alpha$ -helical content of E2-35 correlated with a lower MIC but higher toxicity.<sup>16</sup> Therefore, there is still a need to further optimize AMPs.

## E. Circular Dichroism

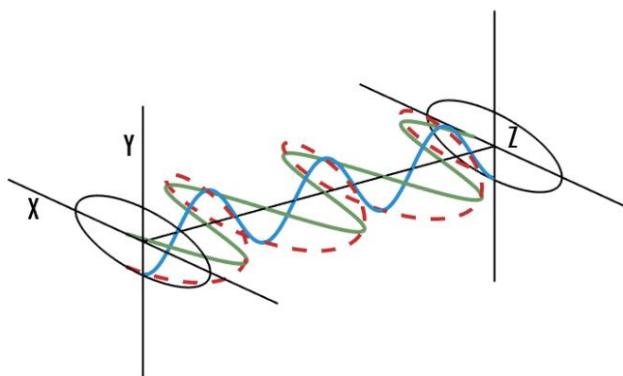
Circular dichroism (CD) is a form of spectroscopy commonly used to determine the relative presence of secondary structures within peptides.<sup>27</sup> Circular dichroism functions through the usage of circularly polarized light, a unique form of polarized light.<sup>29</sup> Normal polarized light, commonly seen in polarized sunglasses, is characterized by the presence of only a single wavelength, usually vertical or horizontal based on the direction of the polarizer.<sup>28,32</sup> Functioning as a sifter, the linear polarizer allows parallel waves of light to pass through unaffected while perpendicular waves are blocked and angled waves are reduced to lesser magnitude parallel light.<sup>28</sup> Importantly, the resulting parallel, linearly polarized light is composed of separate X and Y wave components that then add up to produce the wave.<sup>26,30</sup> To create circularly polarized light, the linearly polarized light is then subjected to a quarter wave-plate which slows down or translates the movement of one of the axes such that when one component intersects their common axis, the other is at a relative maximum or minimum.<sup>28,32</sup> Different wave-plates exist and perform similar functions in the sense that they preserve the speed of one linearly polarized light component while slowing down the other.<sup>31</sup> However, a quarter wave-plate is specifically designed such that the resulting light is circularly polarized; otherwise, a different plate would impose a different slow-down factor and produce elliptical light.<sup>31</sup> Many circular dichroism machines, such as the JASCO-1500 machine that we used for data collection, create circularly polarized light in this manner, otherwise known as phase modulation. Figures 3a-3c and 4 illustrate the aforementioned method of creating circularly polarized light.<sup>23,26,28,29</sup>



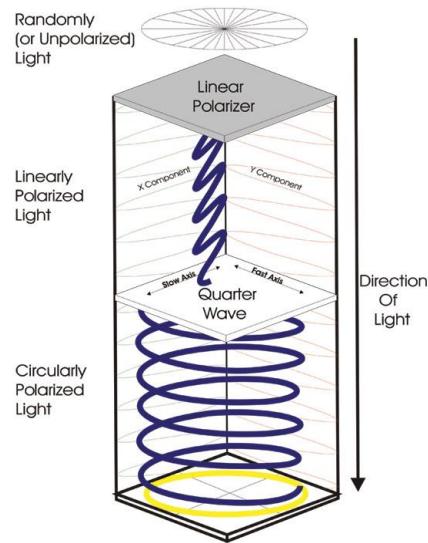
**Figure 3a:** The graph is a 3-D depiction of linearly polarized light (dotted line) with its X and Y components (solid blue and green lines).



**Figure 3b:** This graph depicts circularly polarized light (red dotted line) where the X and Y components (solid blue and green lines) are separated along the Z-axis by a quarter-wavelength.

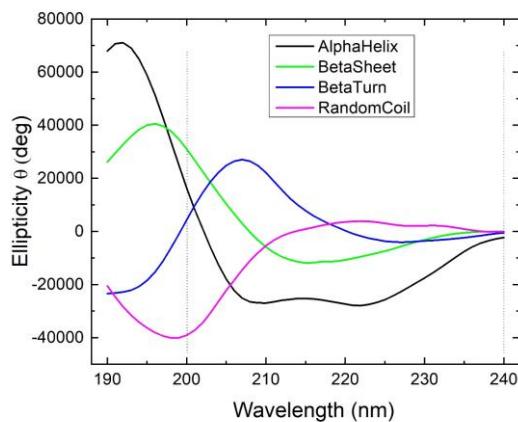


**Figure 3c:** This graph depicts elliptically polarized light (red dotted line) where the Y component of the light (blue solid line) has a lower amplitude than the X component of the light (green solid line), resulting in an overall elliptical rather than circularly shaped light.



**Figure 4:** This diagram demonstrates the entire process of creating circularly polarized light starting from random unpolarized light.

Circular dichroism instruments send circularly polarized light through samples.<sup>26</sup> If the samples are chiral, then they will partially absorb one component of the light, thus reducing its magnitude such that the overall light becomes elliptically polarized.<sup>26</sup> The instrument then measures the levels of ellipticity in a sample as the wavelength changes.<sup>26,31</sup> Wavelengths ranging from 200-250 nanometers are utilized for proteins and peptides as measurable amounts of light are absorbed at those wavelengths.<sup>27</sup> The instrument then measures the levels of ellipticity in a sample, which vary based on the secondary structures present, as the wavelength changes.<sup>24</sup> Proteins made up of a single secondary structure, verified through alternative peptide structural analysis techniques such as X-ray diffraction, produce unique motifs (Fig. 5).<sup>16</sup> These motifs represent ellipticity levels unique to that structure which help determine the structure of an unknown protein.<sup>16</sup>



**Figure 5:** This graph depicts motifs produced by the main secondary structures found in proteins. Specialized data-fitting applies these motifs to data, calculating the relative percentage of secondary structures present in the sample peptide.

Amino acids with aromatic groups, such as phenylalanine, tyrosine, and tryptophan, can also produce their own noticeable motifs if they are present in high volumes.<sup>24</sup> In addition, lipids can contribute some level of ellipticity, though the ones used in this experiment contribute negligible levels. Importantly, the ellipticity is often multiplied by a factor based on the number of amino acids present, providing a metric known as Mean Residue Ellipticity. The circular dichroism graph produced will provide percentages of secondary structures present in the unknown peptide.

## F. Purpose

The purpose of this experiment was to determine the make-up of secondary structures within the cyclic AMP CE-0005B to gain a deeper understanding of the mechanisms involved in peptide-bacterial membrane interactions. Previous studies have established a correlation between  $\alpha$ -helicity in the protein's secondary structure and efficacy in killing bacteria, suggesting that a higher %  $\alpha$ -helicity is associated with an increase in efficacy.<sup>16</sup> These studies, however, admit that the possibility exists for other structures, such as  $\beta$ -sheet,  $\beta$ -turn, and random coil, to be associated with a more favorable bacterial MIC.<sup>16</sup> Beyond determining the structures of this specific peptide, analyzing these structures and their relative effectiveness gives insight into the specific mechanisms of interactions with cell membranes. Further, by gaining a deeper understanding of how AMPs fold when interacting with different types of model membranes and correlating them with their antibiotic effectiveness, we can contribute information to aid the creation of functional AMPs. We can also better predict protein folding patterns for future AMP design.

Our research focused specifically on a cyclic peptide, which contains a circular sequence of bonds. Mika et. al found that cyclic peptides bind more strongly than linear ones to negatively charged membranes, which is useful given bacterial cell membranes are significantly more negatively charged than human cells.<sup>36</sup> The cyclic AMPs are often characterized by a  $\beta$ -sheet structure, rather than an  $\alpha$ -helix. Cyclic peptides also penetrate deeper into the bilayer while the linear peptides remain closer to the surface. This is likely because, upon interaction with the membrane, cyclic peptides tend to fold more, and therefore, are able to perforate the vesicles and cause intracellular leakage.<sup>41</sup>

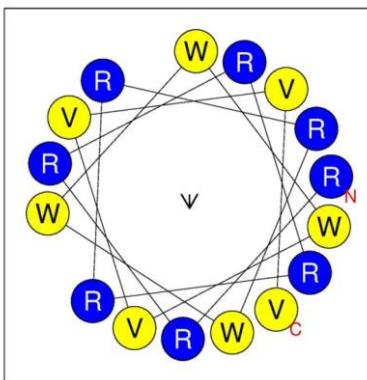
The ratio of peptide to lipid is also important, as different concentrations may have different secondary structures. Our experiment tested for 5 different concentrations and observed the associated changes in secondary structures.

# II. Methods

## A. Sample Preparation

### 1. AMP Synthesis

The CE-0005B antimicrobial peptide samples used were created in Dr. Berthony Deslouches' lab (Fig. 6). These samples consist of arginine (R), tryptophan (W), and valine (V). The full amino acid sequence, RRRRRRRWWVVVV, forms a positively charged hydrophilic region and hydrophobic region. The N-terminus of the peptide is joined to the C-terminus to form a cyclical shape.



**Figure 6: Helical wheel projections depicting the distribution of amino acids in the linear analog of the CE-0005B peptide. This diagram was prepared using the HELIQUEST web server (<https://heliquest.ipmc.cnrs.fr/>). The arrows represent the direction of the hydrophobic moment  $\mu\text{H}$ , which is a measure of helical amphipathicity.**

## 2. Unilamellar Vesicles from Multilamellar Vesicles

Unilamellar vesicles (ULVs) ~600 Å in diameter served as a bacterial model membrane system to simulate the interaction between the CE-0005 peptide and the inner lipid membrane of a Gram-negative bacteria. The lipid components in this lipid model membrane (LMM) are: POPE:POPG:Cardiolipin18:1 (7:2:1 molar ratio); the mixture is referred to as G(-) LMM. To create the ULVs, 100 µL of 20 mg/mL multilamellar vesicle solution in 15 mM phosphate buffered saline (PBS) was pushed through an extruder (Avanti Polar Lipids, Alabaster, AL). The extruder contains a membrane of nucleopores 500 Å in diameter. Passing the lipid mixture through the pores of the extruder membrane results in a model membrane composed of a single lipid bilayer with a final lipid concentration of 15 mg/mL.

In order to make the ULVs, the MLVs were loaded into a 0.25 mL Hamilton syringe, which was inserted into one end of the extruder apparatus. An empty syringe with the plunger set to zero was inserted into the opposite end of the extruder to collect the mixture as it was passed through the membrane. Once the apparatus was assembled, the lipid mixture was passed through the membrane a total of 25 times. The high number of passes ensured the homogeneity of the unilamellar vesicle solution. This procedure was repeated 3 times to ensure a sufficient volume of ULV was obtained for sample preparation. After the appropriate amount of ULV was prepared, concentrated ULVs were added to 3 mL of 10 µM peptide in 15 mM PBS at pH 7. The samples were prepared in varying lipid-to-peptide molar ratios: 5:1, 10:1, 20:1, 30:1, and 50:1. The peptide samples remained at room temperature for 16 hours, sufficient time to embed in the ULV and form secondary structures before data collection occurred.

## B. JASCO 1500 Data Collection

The samples were collected in 3 mL quartz cuvettes. Once the samples were prepared, they were loaded into the JASCO 1500 instrument six at a time to perform CD spectroscopy. The temperature was maintained at 37°C via a Peltier element, simulating the human body, and was verified by using a temperature probe in cell 1. Nitrogen gas was also used at a flow rate between 20 and 25 ft<sup>3</sup>/h in order to deplete excess ozone. The samples were scanned from 200 to 240 nm 20 times with a bandwidth of 1 nm and the results were averaged.

**Table 1: Parameters Used for JASCO Data Collection**

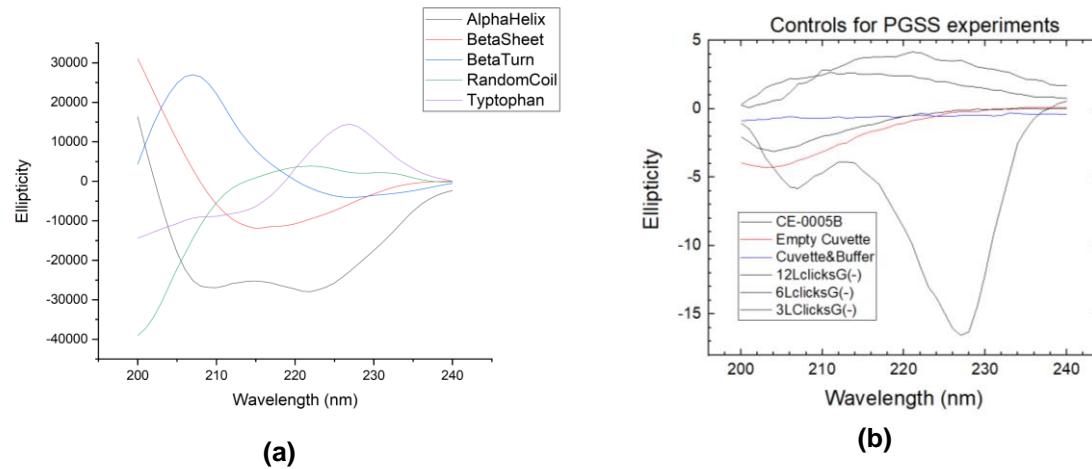
Parameter	Value	Parameter	Value
Temperature	37° C	Scanning Speed	100 nm/min
Wavelength	200-240 nm	Concentration	10 umol/L
Bandwidth	1 nm	Solvent	15mMPBS
Nitrogen Flow Rate	20 and 25 ft <sup>3</sup> /h	Shutter Control	Auto
CD Scale	20 mdeg/0.05 dOD	Baseline Correction	None
Scanning Mode	Continuous	D.I.T	1 s

### C. Data Analysis and Fitting

Data analysis and fitting were performed using OriginPro 2019 Academic (OriginLab, Northampton, Massachusetts). Each data table produced by the JASCO-1500 circular dichroism instrument was inserted into the OriginPro program along with a standardized 'Brahms' dataset which contains the motifs for each secondary structure ( $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn, random coil).<sup>16</sup> To properly scale and align the sample data with the motifs, all motif values along the y-axis were divided by 10,000 while all sample values along the y-axis were raised or lowered such that the ellipticity value at 240 nm is at y=0. If the peptide sample contains large amounts of W, such as the peptide analyzed in this paper, CE-0005B, then a tryptophan 'motif' table was added, scaled/translated to the sample, and then subtracted from the sample. This was to ensure that the new sample data was representative of only the secondary structures present and not dominated by the ellipticity of W. Afterward, a Levenberg-Marquadt fit was performed on the data to provide precise percentages of the secondary structures present within the sample. This algorithm is capable of performing least square curve fits, calculating the ideal  $R^2$  value by combining the Brahms curves to fit the sample ellipticity data.<sup>39</sup> In addition to this original, unweighted fit, weightings were added to certain motifs so as to emphasize their inclusion in the final calculation, creating a range of possible secondary structure percentages with high  $R^2$  values. These fit values are then averaged to create the final secondary structure with standard deviations and  $R^2$  percentages.

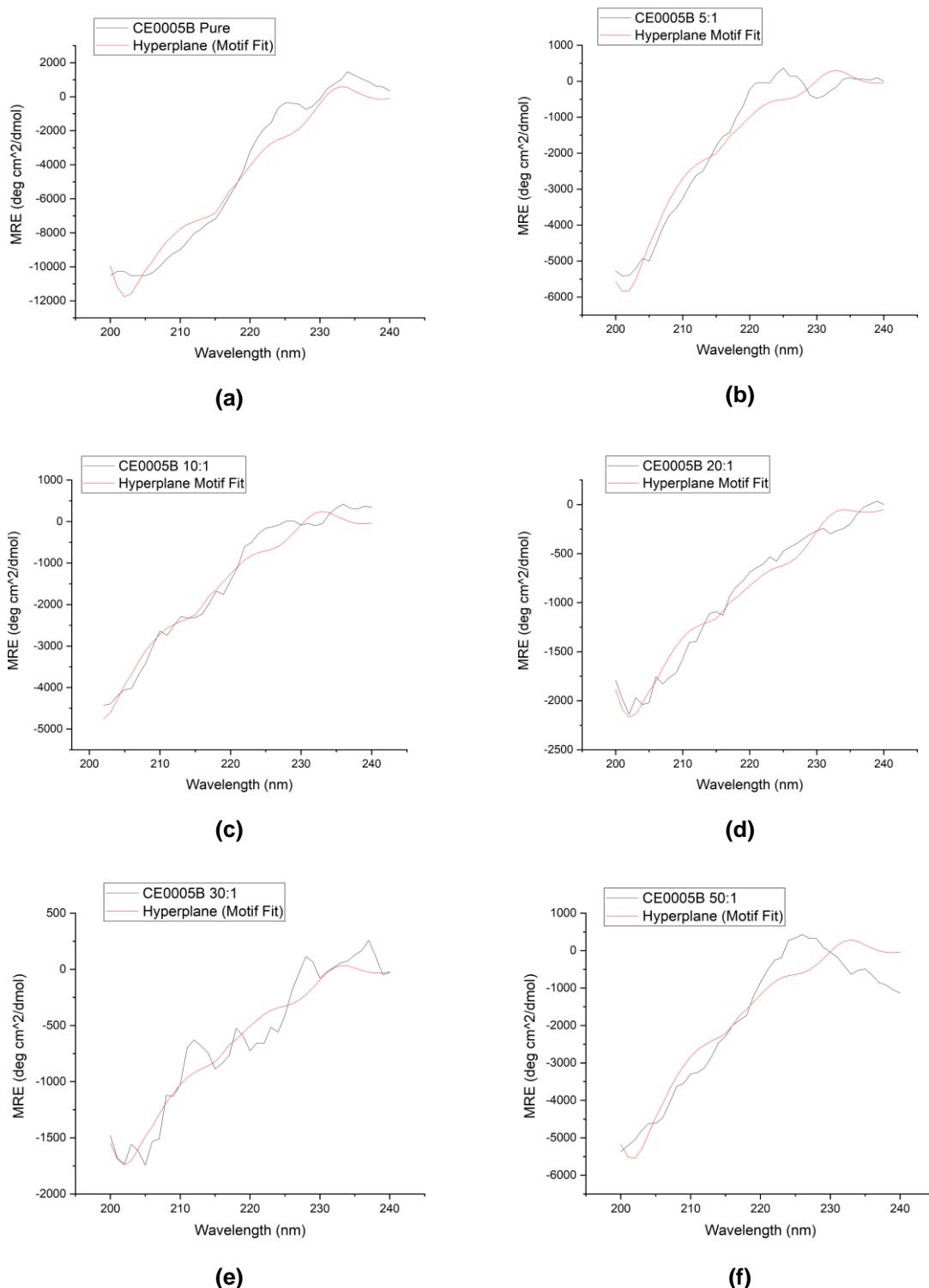
### III. Results

#### A. Data Graphs



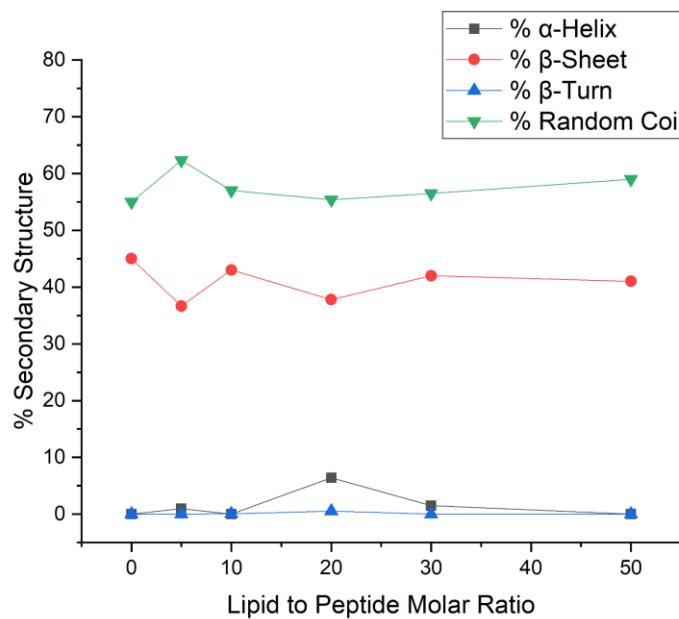
**Figure 7: Ellipticity graphs for motifs and controls.**

(a) Graph displaying motifs for the four secondary structures:  $\alpha$  helix,  $\beta$ -sheet,  $\beta$ -turn, and random coil. These motifs were used to create the hyperplane fit and determine the percentage of each secondary structure. The graph also includes the motif for tryptophan, which was subtracted from the sample before fitting. (b) Graph displaying controls tested, including an empty cuvette, the cuvette and buffer solution with no peptide or lipid, varying amounts of the lipid by itself, and the peptide with no interaction with the lipid. 1 click was 20  $\mu$ L of the ULV suspension. These controls were determined to be insignificant enough to have a negligible effect on our data.



**Figure 8: Comparisons of wavelength (in nm) to mean residue ellipticity (MRE) for different samples of CE-0005B. The black line in a, b, c, e, and f is derived by subtracting tryptophan's ellipticity data from CE-0005B ellipticity data. The red line represents the fitted curve using the four motifs for**

(a) Pure CE-0005B (b) 5:1 lipid-to-peptide ratio (c) 10:1 lipid-to-peptide ratio (d) 20:1 lipid-to-peptide ratio (e) 30:1 lipid-to-peptide ratio (f) 50:1 lipid-to-peptide ratio



**Figure 9: Percent secondary structural motifs vs. lipid-to-peptide molar ratios.**

## B. Fitted Data Lines

**Table 2: Average CE-0005B Secondary Structure Composition**

Antimicrobial Peptide	α-Helix (%)	β-Sheet (%)	β-Turn (%)	Random Coil (%)	Adj. R <sup>2</sup>
Pure CE-0005B	0	45	0	55	0.954
CE-0005B 5:1	1	37	0	62	0.920
CE-0005B 10:1	0	43	0	57	0.966
CE-0005B 20:1	6	38	1	55	0.973
CE-0005B 30:1	2	42	0	57	0.929
CE-0005B 50:1	0	41	0	59	0.889

This table shows the mean percentage compositions of the four secondary structure motifs and the adjusted R-squared values for the six antimicrobial peptide concentrations. The antimicrobial peptides largely consist of  $\beta$ -sheet and random coil structures.

### C. MIC Data

**Table 3: MIC and Toxicity of CE-0005B**

	MIC ( $\mu\text{M}$ )						% Toxicity	
	PA	AB	KP	<i>E. coli</i>	<i>Enterobacter</i>	Average	RBC	WBC
<b>CE-0005B</b>	6.1	1.1	6.7	2.5	3.2	<b><math>3.9 \pm 1.1</math></b>	0.00	0.00
<b>Colistin</b>	8.4	0.5	0.7	4.3	12.1	<b><math>5.2 \pm 4.3</math></b>	--	--
<b>Tobramycin</b>	32.0	32.0	2.1	28.0	24.5	<b><math>23.7 \pm 3.4</math></b>	--	--

Table 3 depicts the minimum inhibitory concentrations (MIC) of CE-0005, Colistin, and Tobramycin against various Gram-negative [G(-)] bacteria. MIC represents the minimum concentration of substance necessary to stop bacterial growth. The G(-) bacterial strains are *Pseudomonas aeruginosa* (PA), *Acinetobacter baumannii* (AB), *Klebsiella pneumoniae* (KP), *Escherichia coli*, and *Enterobacter*. Colistin is an established antimicrobial peptide used in worst-case scenarios while Tobramycin is a conventional antibiotic. RBC and WBC represent human red and white blood cell lysis/toxicity percentages with 32 and 16  $\mu\text{M}$  of peptide respectively. Data is representative of 2-3 experimental trials while the bacterial isolates were obtained from UPMC.

## IV. Discussion

### A. Data Analysis

Our data indicate that CE-0005B has primarily  $\beta$ -sheet and random coil secondary structures across all lipid-to-peptide molar ratios. According to Table 3, the MIC of CE-0005B is less than that of Colistin, a previously established AMP with known  $\alpha$ -helices. This suggests that  $\beta$ -sheet and random coil secondary structures may be the cause of the higher antimicrobial effects, though the difference in the MICs is not statistically significant.

In addition, the % red blood cell (RBC) and white blood cell (WBC) lysis are 0%, indicating that CE-0005B displays no hemolytic activity on mammalian red and white blood cells and thus shows no signs of toxicity. This is a significant finding because other antimicrobial peptides, such as colistin, show high toxicity.

The data from Fig. 8a displaying the pure peptide with no added lipid show a similar structure to the peptide when interacting with the lipids. Often, many antimicrobial peptides are random coil in aqueous solution and only adopt a secondary structure upon association with the membrane. We now show that cyclic AMPs, can adopt a  $\beta$ -sheet and random coil secondary structure even in aqueous solution, independent of membrane interaction. For example, gramicidin S has a  $\beta$ -sheet structure in pure water.<sup>36</sup> This explains

why the pure peptide still exhibited similar secondary structures without interactions. This finding could be useful as this unique property of cyclic peptides confers some key advantages when compared to linear peptides. Mika et. al explain that small linear peptides often lack a well-defined structure, which makes them highly flexible and significantly more soluble in water. The increased rigidity of cyclic peptides is advantageous by pre-organizing polypeptides for receptor binding. The aqueous structure of a cyclic peptide also has to change less than a random peptide structure in water to bind to a protein.

Dr. Mitra et al. discovered that  $\alpha$ -helical AMPs tend to exhibit higher efficacy; however, our data deviate from these findings.<sup>16</sup> Mika et al.'s findings validate our results, as they found that the cyclical AMP structures involved in forming  $\beta$ -sheets penetrate deeper into the bacterial membrane, resulting in increased bacterial cell lysis compared to linear AMPs. Additionally, low hemolytic activity was observed in mammalian cells when cyclic peptides were introduced, further supporting our findings.

## B. Limitations

Our first limitation involved the positive control group. During testing, a sample of the helical peptide E2-35 from our original control group did not appear to have any helicity. This occurred on multiple occasions, so to rule out the possibility of extraneous factors affecting the control or the sample, a set of controls was run, including an empty cuvette, cuvette and buffer solution, pure peptide, and varying amounts of lipid. Figure 8b depicts these controls. The best explanation for the nonhelical structure of E2-35 is that the peptide is more than a year old, and while it was stored in a freezer at -20°C, denaturation and degradation were extremely likely. Furthermore, the peptide contains arginine and tryptophan which are extremely reactive groups and are likely to oxidize and react, thus affecting structure. A more representative positive helical control would have been beneficial to further validate the results. Nevertheless, the additional controls were sufficient to verify that the results obtained for the CE-0005B samples were proper representations of their secondary structure.

The second limitation involved further validation of our results. While circular dichroism provides a relatively accurate view into the secondary structure of the peptide, it doesn't precisely describe the atomic-level interactions between the peptide and lipids. As outlined in the introduction, X-ray diffuse scattering provides data about the location of AMPs, membrane stiffness, and lipid chain order. Combining these data with CD data provides a precise understanding of the AMP's mechanism of action. Such an understanding is critical for long-term and clinical applications of AMPs, though CD data forms the necessary starting point.

## V. Conclusions

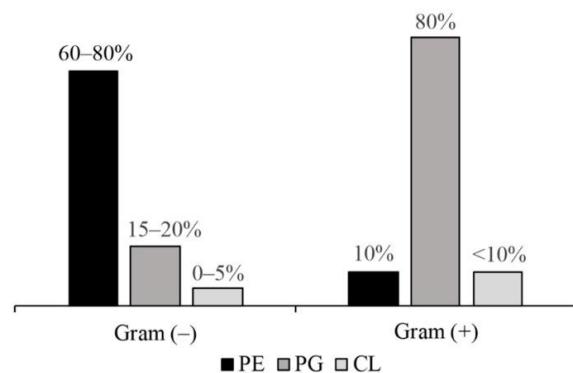
Over the years, the increasing emergence of antimicrobial-resistant pathogens has contributed to a decline in the efficacy of established antibiotic treatments. AMPs have shown promising results in addressing antimicrobial resistance; however, to ensure the safety of AMP use as an alternative to antibiotics, researchers are attempting to identify their structures and mechanisms of action. The purpose of this study was to determine the secondary structure of the cyclic AMP CE-0005B and understand the mechanisms involved with bacterial cell destruction. Previous studies have established correlations between  $\alpha$ -helicity in a protein's secondary structure and efficacy in killing bacteria. These studies also admit the possibility for other structures, such as  $\beta$ -sheet,  $\beta$ -turn, and random coil, to be correlated with the level of bacterial minimum inhibitory concentration (MIC). CE-0005B AMP samples were created prior to the experiment in Dr. Berthony Deslouches' lab. Gram-negative unilamellar vesicles (ULVs) served as a model membrane system to simulate interaction between CE-0005 peptide and lipid membrane of Gram-negative bacteria. The samples were loaded into the JASCO J-1500 instrument to perform CD spectroscopy. Data analysis using the Levenberg-Marquadt fit was performed using OriginPro 2019, producing nonlinear curve fits to

create a range of possible secondary structure percentages with high Adjusted R<sup>2</sup> values. CE-0005B was determined to be composed of higher percentages of β-sheet and random coil secondary structures.

Cyclic peptides, such as CE-0005B, have demonstrated the strength of their stable structure in binding small targets. Identifying the structures of these cyclic peptides can provide insights and inspiration for designing new AMPs.<sup>36</sup> Cationicity and amphipathicity determine the readiness of cyclic AMPs for clinical applications.<sup>14</sup> For example, amphipathicity determines the ability of AMPs to insert into bacterial cell membranes and form hydrophobic channels. While these factors may affect the clinical implementation of AMPs, identifying structures within CE-0005B leads to a promising future of alternative antibiotic treatment.

## VI. Future Directions

While data from the lipid model membranes mimic that of Gram-negative bacteria, the exploration of interactions between antimicrobial peptides and lipid membranes mimicking Gram-positive bacteria and eukaryotic cells should also be studied due to different molecular properties. Gram-positive bacteria have a greater negative charge density due to a high concentration of phosphatidylglycerol (PG), facilitating AMP action. Figure 10 shows the differing phospholipid compositions of Gram-negative and Gram-positive bacterial membranes.<sup>39</sup> AMPs need to diffuse across the peptidoglycan matrix before acting on the cytoplasmic membrane of Gram-positive bacteria, while killing Gram-negative bacteria consists of perforating both outer and cytoplasmic membranes.<sup>14</sup> Gram-negative peptidoglycan is a few nanometers thick, containing few layers while Gram-positive peptidoglycan is 30-100 nanometers thick, containing many layers.<sup>40</sup> Since the penetration mechanisms of Gram-positive and Gram-negative bacteria vary, further exploration of CE-0005B's action on Gram-positive membranes would provide a better understanding of the role of specific secondary structures.

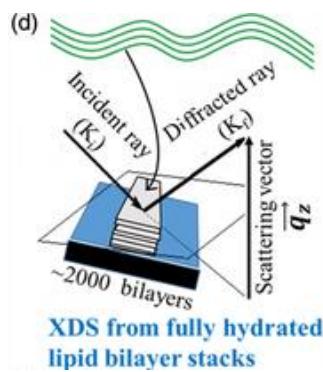


**Figure 10:** The two graphs depict the phospholipid compositions of Gram-negative and Gram-positive bacterial membranes. The percentages of phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL) are shown.

Cyclic peptides fold at the membrane interface and adopt a β-sheet structure characterized by two turns, aligning with the results of the circular dichroism data.<sup>36</sup> However, the CE-0005B peptide retained a β-sheet and random coil compositions throughout all lipid and peptide ratios. Further studies on other cyclic peptides would provide insight into whether a β-sheet and random coil composition is a specific pattern.

X-ray diffuse scattering (XDS) data would confirm the membrane elasticity and interaction mechanisms at play. After circular dichroism determines the secondary structures of the AMPs, X-ray diffuse scattering is used to determine the elastic modulus and lipid chain order to better understand the interactions between the peptide and the lipid. Resulting, a correlation can be made between the secondary structure of the

peptide and the MIC values and the elastic modulus of the lipid membrane. Low-angle XDS is used to determine the elastic modulus ( $K_c$ ) which measures the amount of energy needed to bend a single layer of membrane and the compressibility modulus ( $B$ ) which represents the bending and interaction of multiple layers within the membrane. The higher the  $K_c$  value, the stiffer the membrane and the lower the  $K_c$  value, the softer the membrane. Low angle XDS also shows the location of the peptide in the lipid bilayer. Wide angle XDS is used to determine the lipid chain order ( $S_{x\text{-ray}}$ ) — higher values signify ordered lipid acyl chains and lower values signify disordered lipid acyl chains. The various uses of X-ray diffuse scattering work to identify mechanisms contributing to the MIC values of the peptide. An existing example is an X-ray diffuse scattering study that confirmed a change in membrane elasticity and lipid chain ordering of lipid membranes in the presence of colistin, suggesting an action involving lipid domains. This method of analysis also identified changes in elasticity similar to colistin produced by WLBU2 and D8-WLBU2, a stereoisomer. Figure 11 shows the X-ray diffuse scattering setup to obtain the structure of the lipid LMMs in the presence of antimicrobial peptides.<sup>16</sup> The experimental methods of this paper utilized circular dichroism to analyze the secondary structures of antimicrobial peptides — E235 and CE-0005B — utilizing UV light. CD is complementary to XDS.



**Figure 11:** This shows the X-ray diffuse scattering setup from fully hydrated lipid bilayer stacks on silicon wafers for obtaining material properties, structure of the lipid LMMs in the presence of antimicrobial peptides, and lipid chain order.

Due to similar secondary structure compositions for both the aqueous and lipid environments, the synthesis of this data with computer molecular modeling visualization should be considered in the study of the underlying similarity of the peptide interactions in multiple environments.

## VII. Acknowledgments

First and foremost, we would like to thank the Pennsylvania Governor's School of Science (PGSS) Alumni Campaign for providing us with the opportunity to conduct this research. The AMP project was funded through Dr. Stephanie Tristram-Nagle's collaboration with Dr. Toni Deslouches, who synthesized the AMPs, at the University of Pittsburgh. These investigators are funded through NSF and NIH. We are appreciative of Carnegie Mellon University and Mellon College of Science for allowing us access to their facilities throughout the program. We would like to extend our thanks to our team project advisor Dr. Stephanie Tristram-Nagle and lab members Dr. Saheli Mitra, Mio Chen, and Fran Stedman for their guidance throughout our research and the project presentation process. We would also like to express our gratitude to our teaching assistant Nikki Holmes for her continued support. A special thanks to Dr. Barry Luokkala, Melissa Lessure, Cameron Blackwood-Short, and the PGSS Faculty and Staff for making PGSS an incredible and enriching experience.

## VIII. References

1. Ventola, C. L. (2015). The Antibiotic Resistance crisis: Part 1: Causes and Threats. *P & T : A Peer-Reviewed Journal for Formulary Management*, 40(4), 277–283.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/>
2. Society, M. (2023). *The history of antibiotics*. Microbiologysociety.org.  
<https://microbiologysociety.org/membership/membership-resources/outreach-resources/antibiotics-unearthed/antibiotics-and-antibiotic-resistance/the-history-of-antibiotics.html>
3. *Outpatient Antibiotic Prescriptions — United States, 2020 | Antibiotic Use | CDC*. (2022, October 24). [Www.cdc.gov.](https://www.cdc.gov)
4. *Nearly a quarter of antibiotic prescriptions for children and adults may be unnecessary*. (n.d.).  
<https://www.michiganmedicine.org/health-lab/nearly-quarter-antibiotic-prescriptions-may-be-unnecessary>
5. World Health Organization. (2021, November 17). *Antimicrobial resistance*. Who.int; World Health Organization: WHO. <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
6. *Mutations and selection – Antibiotic resistance*. (n.d.). ReAct. Retrieved August 3, 2023, from <https://www.reactgroup.org/toolbox/understand/antibiotic-resistance/mutation-and-selection/#:~:text=If%20we%20were%20to%20treat>
7. Review On Antimicrobial Resistance, & Wellcome Trust (London, England. (2016). *Tackling drug-resistant infections globally : final report and recommendations*. Review On Antimicrobial Resistance, May.
8. Centers For Disease Control And Prevention (U.S. (2019). *Antibiotic resistance threats in the United States, 2019*. U.S. Department Of Health And Human Services, Centres For Disease Control And Prevention.
9. Rima, M., Rima, M., Fajloun, Z., Sabatier, J.-M., Bechinger, B., & Naas, T. (2021). Antimicrobial Peptides: A Potent Alternative to Antibiotics. *Antibiotics (Basel, Switzerland)*, 10(9), 1095.  
<https://doi.org/10.3390/antibiotics10091095>
10. Huan, Y., Kong, Q., Mou, H., & Yi, H. (2020). Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Frontiers in Microbiology*, 11.  
<https://doi.org/10.3389/fmicb.2020.582779>
11. Rehman, I., Farooq, M., & Botelho, S. (2022). Biochemistry, Secondary Protein Structure. In *StatPearls*. StatPearls Publishing.
12. *Secondary Structure - Beta turns & Random coils*. (n.d.). Comis.med.uvm.edu.  
[https://comis.med.uvm.edu/VIC/coursefiles/MD540/MD540-Protein\\_Organization\\_10400\\_574581210/Protein-org/Protein\\_Organization5.html#:~:text=and%20Random%20Coils-](https://comis.med.uvm.edu/VIC/coursefiles/MD540/MD540-Protein_Organization_10400_574581210/Protein-org/Protein_Organization5.html#:~:text=and%20Random%20Coils-)

13. Zhang, Q. Y., Yan, Z. B., Meng, Y. M., Hong, X. Y., Shao, G., Ma, J. J., Cheng, X. R., Liu, J., Kang, J., & Fu, C. Y. (2021). Antimicrobial peptides: mechanism of action, activity and clinical potential. *Military Medical Research*, 8(1), 48. <https://doi.org/10.1186/s40779-021-00343-2>
14. Li, J., Koh, J. J., Liu, S., Lakshminarayanan, R., Verma, C. S., & Beuerman, R. W. (2017). Membrane Active Antimicrobial Peptides: Translating Mechanistic Insights to Design. *Frontiers in neuroscience*, 11, 73. <https://doi.org/10.3389/fnins.2017.00073>
15. Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D. H., & He, Q. (2019). The antimicrobial peptides and their potential clinical applications. *American journal of translational research*, 11(7), 3919–3931.
16. Mitra, S., Coopershlyak, M., Li, Y., Chandersekhar, B., Koenig, R., Chen, M. T., Evans, B., Heinrich, F., Deslouches, B., & Tristram-Nagle, S. (2023). Novel Helical Trp- and Arg-Rich Antimicrobial Peptides Locate Near Membrane Surfaces and Rigidify Lipid Model Membranes. *Advanced nanobiomed research*, 3(5), 2300013. <https://doi.org/10.1002/anbr.202300013>
17. Wimley W. C. (2010). Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS chemical biology*, 5(10), 905–917. <https://doi.org/10.1021/cb1001558>
18. Jäkel, C. E., Meschenmoser, K., Kim, Y., Weiher, H., & Schmidt-Wolf, I. G. (2012). Efficacy of a proapoptotic peptide towards cancer cells. In vivo (Athens, Greece), 26(3), 419–426.
19. Tan, P., Wu, C., Tang, Q., Wang, T., Zhou, C., Ding, Y., Fu, H., Xu, S., Feng, Y., Zhang, Y., Dai, Q., & Ma, X. (2023). pH-Triggered Size-Transformable and Bioactivity-Switchable Self-Assembling Chimeric Peptide Nanoassemblies for Combating Drug-Resistant Bacteria and Biofilms. *Advanced materials (Deerfield Beach, Fla.)*, 35(29), e2210766. <https://doi.org/10.1002/adma.202210766>
20. Deslouches, B., Islam, K., Craig, J. K., Paranjape, S. M., Montelaro, R. C., & Mietzner, T. A. (2005). Activity of the de novo engineered antimicrobial peptide WLBU2 against *Pseudomonas aeruginosa* in human serum and whole blood: implications for systemic applications. *Antimicrobial agents and chemotherapy*, 49(8), 3208–3216. <https://doi.org/10.1128/AAC.49.8.3208-3216.2005>
21. Vogel, H. J., Schibli, D. J., Jing, W., Lohmeier-Vogel, E. M., Epand, R. F., & Epand, R. M. (2002). Towards a structure-function analysis of bovine lactoferricin and related tryptophan- and arginine-containing peptides. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 80(1), 49–63. <https://doi.org/10.1139/o01-213>
22. Mandell, J. B., A Koch, J., Deslouches, B., & Urich, K. L. (2020). Direct antimicrobial activity of cationic amphipathic peptide WLBU2 against *Staphylococcus aureus* biofilms is enhanced in physiologic buffered saline. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, 38(12), 2657–2663. <https://doi.org/10.1002/jor.24765>
23. Tristam-Nagle, S. (2023, July). PGSS. <https://docs.google.com/presentation/d/1J4OKXQsE7DGV7fbtzTXPJtybuKi71GBp/edit#slide=id.p6>

24. 4.2.1 *Circular dichroism spectroscopy*. (n.d.). [www.cryst.bbk.ac.uk](http://www.cryst.bbk.ac.uk). Retrieved August 3, 2023, from [https://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531\\_21.html#HEADING27](https://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531_21.html#HEADING27)
25. *Beyond Conventional Imaging: Sony's Polarized Sensor - Polarsens | LUCID Vision Labs*. (n.d.). <https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/>
26. *Circular Dichroism*. (2013, October 2). Chemistry LibreTexts. [https://chem.libretexts.org/Bookshelves/Physical\\_and\\_Theoretical\\_Chemistry\\_Textbook\\_Maps/Supplemental\\_Modules\\_\(Physical\\_and\\_Theoretical\\_Chemistry\)/Spectroscopy/Electronic\\_Spectroscopy/Circular\\_Dichroism#:~:text=Circular%20Dichroism%20is%20an](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Electronic_Spectroscopy/Circular_Dichroism#:~:text=Circular%20Dichroism%20is%20an)
27. *Circular Polarization*. (n.d.). [Www.youtube.com](https://www.youtube.com/watch?v=ycY2mUZHS84). <https://www.youtube.com/watch?v=ycY2mUZHS84>
28. *Create Circularly Polarized Light Using a Quarter-Wave Plate (QWP) | Thorlabs Insights*. (n.d.). [Www.youtube.com](https://www.youtube.com/watch?v=P0asuzX4x-Q). Retrieved August 3, 2023, from <https://www.youtube.com/watch?v=P0asuzX4x-Q>
29. *Circular Polarizer | How Does It Work? - American Polarizers*. (2019, February 18). American Polarizers, Inc. <https://www.apioptics.com/about-api/api-blog/api-news/how-circular-polarization-works/>
30. *Polarization*. (n.d.). Electron6.Phys.utk.edu. Retrieved August 3, 2023, from <http://electron6.phys.utk.edu/optics421/modules/m7/polarization.htm>
31. Circular Dichroism J-1000 Series Spectrometers Performance Innovation Reliability. (n.d.). Retrieved August 4, 2023, from <https://jascoinc.com/wp-content/uploads/2018/11/CD-Brochure-11.18.pdf>
32. Paschotta, D. R. (n.d.). Phase Modulators. [Www.rp-Photonics.com](https://www.rp-photonics.com/phase_modulators.html). [https://www.rp-photonics.com/phase\\_modulators.html](https://www.rp-photonics.com/phase_modulators.html)
33. Small Angle X-ray Scattering (SAXS). (n.d.). Retrieved August 4, 2023, from <https://www.eng.uc.edu/~beaucag/Classes/Analysis/Chapter8.pdf>
34. Van Benschoten, A. H., Liu, L., Gonzalez, A., Brewster, A. S., Sauter, N. K., Fraser, J. S., & Wall, M. E. (2016). Measuring and modeling diffuse scattering in protein X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15), 4069–4074. <https://doi.org/10.1073/pnas.1524048113>
35. Dissertation, R. (n.d.). X-RAY THERMAL DIFFUSE SCATTERING AND ITS STUDIES OF LATTICE DYNAMICS. Retrieved August 4, 2023, from <https://core.ac.uk/download/pdf/4827061.pdf>
36. Mika, J. T., Moiset, G., Cirac, A. D., Feliu, L., Eduard Bardají, Planas, M., Sengupta, D., Marrink, S. J., & Poolman, B. (2011). *Structural basis for the enhanced activity of cyclic antimicrobial peptides: The case of BPC194*. 1808(9), 2197–2205. <https://doi.org/10.1016/j.bbamem.2011.05.001>
37. Gram-positive and gram-negative: What is the difference? (2022, September 29). [www.medicalnewstoday.com](http://www.medicalnewstoday.com).

38. Gavin, H. (2022). *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. <https://people.duke.edu/~hpgavin/ExperimentalSystems/lm.pdf>
39. Barreto-Santamaría, A., Arévalo-Pinzón, G., Patarroyo, M. A., & Patarroyo, M. E. (2021). How to Combat Gram-Negative Bacteria Using Antimicrobial Peptides: A Challenge or an Unattainable Goal?. *Antibiotics (Basel, Switzerland)*, 10(12), 1499. <https://doi.org/10.3390/antibiotics10121499>
40. Silhavy, T. J., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harbor perspectives in biology*, 2(5), a000414. <https://doi.org/10.1101/cshperspect.a000414>
41. Malde, A. K., Hill, T. A., Iyer, A., & Fairlie, D. P. (2019). Crystal structures of protein-bound cyclic peptides. *Chemical Reviews*, 119(17), 9861–9914. <https://doi.org/10.1021/acs.chemrev.8b00807>

**CHEMISTRY  
TEAM PROJECTS**



# **Analysis of Atmospheric Trends and Climate Change Impacts Using Pittsburgh Weather Balloon Data**

Nidhi Ram, Krishna Ravi, Ryan Sajwan, Rachael Saldubehere

## **Abstract**

This study observed the local impacts of climate change using historic radiosonde data from a Pittsburgh upper air station. Meteorological profiles, including temperature, atmospheric pressure, wind speed, and relative humidity, and meteorological indices from the sounding columns of the weather balloons, collected from the University of Wyoming Upper Air Database, were processed, visualized, and analyzed to understand temporal trends over the study period from 1990–2023. Median surface and tropopause trends for the dataset, split into five-year periods, were analyzed and compared to examine the relationship between local atmospheric layers and current knowledge of global trends due to the impacts of climate change. Differences between local and global trends were found, as several proven global trends, including relative humidity, mean mixing layer mixing ratio, and tropopause height, were not observed as clearly on a local scale. Not all impacts of climate change are observable regionally as it is a global phenomenon, but several trends, including wind speed and pressure for both the surface and tropopause, are distinguishable on a local scale in Pittsburgh.

## **I. Introduction**

### **A. The Earth's Atmosphere**

The atmosphere is divided into five main layers.<sup>1</sup> The layer closest to the Earth's surface is the troposphere, which extends about 12 kilometers upwards.<sup>1</sup> This layer contains 99% of all atmospheric water vapor, and almost all of Earth's weather and clouds are found here.<sup>1</sup> Temperatures decrease with altitude in this layer because the interaction of solar radiation with the Earth's surface generates heat located in the troposphere,<sup>1</sup> unlike layers located farther from the Earth's surface where solar radiation primarily controls thermodynamics.<sup>2</sup> As solar radiation is absorbed and re-radiated back towards the atmosphere, the air molecules located further away from the Earth's surface contain less heat than molecules closer to the surface, creating a negative correlation between temperature and altitude in the troposphere.<sup>2</sup> The layer above the troposphere is the stratosphere, which extends to around 50 kilometers above the Earth's surface.<sup>1</sup> The correlation between temperature and altitude in the stratosphere is opposite to that of the troposphere.<sup>1</sup> This is because the ozone layer is located in the stratosphere and is responsible for absorbing harmful ultraviolet (UV) radiation from the sun,<sup>1</sup> which generates heat.<sup>2</sup> The lapse rate, the ratio of the rate of change in temperature to the rate of change in height, for the troposphere and the stratosphere are opposite in sign.<sup>2</sup> The calculation for the lapse rate is shown below in Equation 1. A positive lapse rate of the troposphere indicates that temperature decreases with height, while a negative lapse rate of the stratosphere indicates that temperature increases with height.<sup>2</sup> The boundary sublayer between the troposphere and the stratosphere is called the tropopause.<sup>2</sup> Starting above the stratosphere, the three other layers of the atmosphere include the mesosphere, thermosphere, and exosphere.<sup>1</sup> However, radiosonde data is rarely collected from these layers, as weather balloons cannot ascend past the stratosphere.<sup>3</sup> To measure atmospheric elements above the stratosphere, such as in the mesosphere, sounding rockets and rocket-propelled aircrafts are used to reach the mesosphere<sup>1</sup>.

$$\text{Lapse Rate} = -\frac{dT}{dz}$$

Equation 1: Lapse Rate Formula

## B. Climate Change

One fundamental factor of climate change is the amplification of the greenhouse effect. The process itself is natural and essential to sustaining life on Earth, as surface temperatures would drop 33°C without it.<sup>5</sup> The greenhouse effect is caused by greenhouse gasses (GHGs) in the atmosphere, including carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), industrial gasses, ozone (O<sub>3</sub>), and water vapor.<sup>6</sup> Although greenhouse gasses only comprise 0.43% of air molecules in the atmosphere,<sup>7</sup> they contribute to a positive feedback loop in the climate that can considerably warm the Earth.<sup>8</sup> Some solar radiation passes through the Earth's atmosphere and is absorbed by the Earth's surface. Because the Earth is cooler than the sun, solar energy is radiated back into the atmosphere as infrared radiation (IR), a lower energy form of radiation than visible light.<sup>9</sup> The infrared radiation is then re-emitted by the GHGs, sending energy back towards the Earth, out to space, and even to other GHG molecules.<sup>9</sup> The absorption of IR prevents energy from escaping the atmosphere, which warms the planet.<sup>5</sup> Deforestation, burning of fossil fuels, and other human activities, which steadily emit greenhouse gasses into the atmosphere, have significantly amplified this natural cycle, however.<sup>9</sup>

## C. Weather Balloons

Radiosondes measure state parameters in the atmosphere, including relative humidity, temperature, and pressure.<sup>10</sup> This data is often used for weather forecasting and climate change research.<sup>12</sup> Wind speed and direction can also be measured using GPS tracking.<sup>10</sup> This data is collected using weather balloons which contain the radiosonde that collects the data, a balloon filled with either helium or hydrogen, and a small parachute.<sup>10</sup> A radio transmitter linked to sensors on the radiosonde transmits the newly acquired data to a transmitter on the ground.<sup>10</sup> Weather balloons are launched twice daily from stations worldwide at 0 and 12 Zulu time,<sup>10</sup> which correlates to 7 or 8 pm and 7 or 8 am in Pittsburgh, PA, based on daylight saving time.<sup>11</sup> There are approximately 1300 stations worldwide, referred to as upper-air stations.<sup>10</sup> The weather balloons can travel 200 kilometers laterally and reach their height limit at around 30–35 kilometers in the stratosphere.<sup>10</sup> Weather balloons burst above this height because of the low pressure of the balloon's environment.<sup>12</sup>

## D. Tropopause Height

The boundary between the troposphere and the stratosphere is called the tropopause.<sup>26</sup> The tropopause height is defined as the lowest level where the lapse rate decreases to 2°C/km or less.<sup>2</sup> The tropopause height can change based on temperature changes in the troposphere or the stratosphere, making it an excellent indicator of climate change.<sup>13</sup> Although natural events, such as solar irradiance and volcanic aerosols<sup>26</sup>, can influence the height of the tropopause, two main factors generally account for these changes: tropospheric warming and stratospheric cooling.<sup>13</sup>

The continuous release of GHGs causes tropospheric warming.<sup>13</sup> GHGs absorb and re-emit IR radiation that had been converted from solar radiation in all layers of the atmosphere; however, GHGs can trap more heat in the troposphere due to the density of the air compared to upper atmospheric layers.<sup>14</sup> In the

troposphere, GHG molecules transfer energy generated from the absorption of IR to other nearby air molecules and heat the air.<sup>14</sup> Meanwhile, the increased concentration of CO<sub>2</sub> in the upper layers of the atmosphere has the opposite effect on temperature in the stratosphere. The stratosphere is less dense than the troposphere. Because there are fewer air molecules present, the majority of the energy produced from the absorption of IR by CO<sub>2</sub> is lost to space.<sup>14</sup> The emission of ozone-depleting substances also contribute to the cooling of the stratosphere by destroying parts of the ozone layer, which counters adiabatic cooling within the stratosphere.<sup>14</sup> Since the 1980s, the stratosphere has shrunk 400 meters and is expected to contract another 900 meters by the year 2080.<sup>15</sup> The temperature trends observed in the troposphere and stratosphere change the thermodynamic structure of both sides of the tropopause and therefore impact the location of the tropopause.<sup>13</sup> A study conducted in 2021 found that the tropopause grew in height at a rate of 50–60 m/decade from 2001–2020, with a comparable rate from 1980–2000<sup>13</sup>. Additionally, tropospheric warming was a stronger contributing factor than stratospheric cooling from 2000–2020, which could be a result of the increase in greenhouse gas emissions and decrease in emissions of ozone-depleting substances over the years 2001–2020.<sup>13</sup>

## E. Water Vapor and Humidity

Water vapor is the most abundant GHG.<sup>8</sup> It is also part of a positive feedback loop, which contributes to the warming of the Earth to such an extent that doubling the amount of water vapor has a stronger warming effect than doubling carbon dioxide.<sup>8</sup> The loop begins when GHGs are emitted into the atmosphere and start to warm up the air because of the greenhouse effect.<sup>8</sup> The warmed atmosphere can hold more water vapor due to the Clausius-Clapeyron relationship,<sup>22</sup> with a 7% increase in water holding capacity per degree Celsius increase.<sup>8</sup> Global warming also increases evaporation, which then creates an increase in precipitation.<sup>8</sup> More water vapor accumulates in the atmosphere, and more IR radiation is absorbed because water vapor is a GHG.<sup>8</sup> Half of the greenhouse effect is caused by this positive feedback loop with water vapor.<sup>8</sup> One significant measurement collected by weather balloons is humidity, which is measurable in two metrics: absolute and relative humidity. Absolute humidity is the actual concentration of water vapor in the air,<sup>16</sup> whereas relative humidity is the concentration of water vapor relative to the air temperature.<sup>16</sup> Relative humidity, however, is expressed as a percentage of the amount of water vapor present out of the maximum amount of water vapor the air can hold at a specific temperature.<sup>16</sup> For example, a parcel of air with a set absolute humidity will have a higher relative humidity in cooler temperatures and a lower relative humidity in warmer temperatures.<sup>16</sup> The mean mixing layer mixing ratio (MLMR) was used to measure humidity because radiosondes do not measure absolute humidity. MLMR is the ratio of the mass of water vapor present to the mass of the air in the mixing layer of the atmosphere.<sup>2</sup> The atmospheric mixed layer is characterized by intense turbulence from strong winds resulting in the vertical mixing of air pollutants, tracers, potential temperature, and wind speed.<sup>17, 18, 20</sup>

## F. CAPE and CIN

Convective Available Potential Energy, or CAPE, is a metric indicating the potential for a thunderstorm based on the amount of convective energy available in the atmosphere.<sup>20</sup> CAPE measures the amount of positive buoyancy a rising parcel of air has in a specific environment (Fig. 1).<sup>21</sup> Although there is a specified range describing the relative strength of CAPE values, these ranges can shift based on the time and location of when CAPE is measured.<sup>20</sup> CAPE values are relative to the current environment, so a CAPE value considered weak in the spring could be strong in the winter.<sup>20</sup>

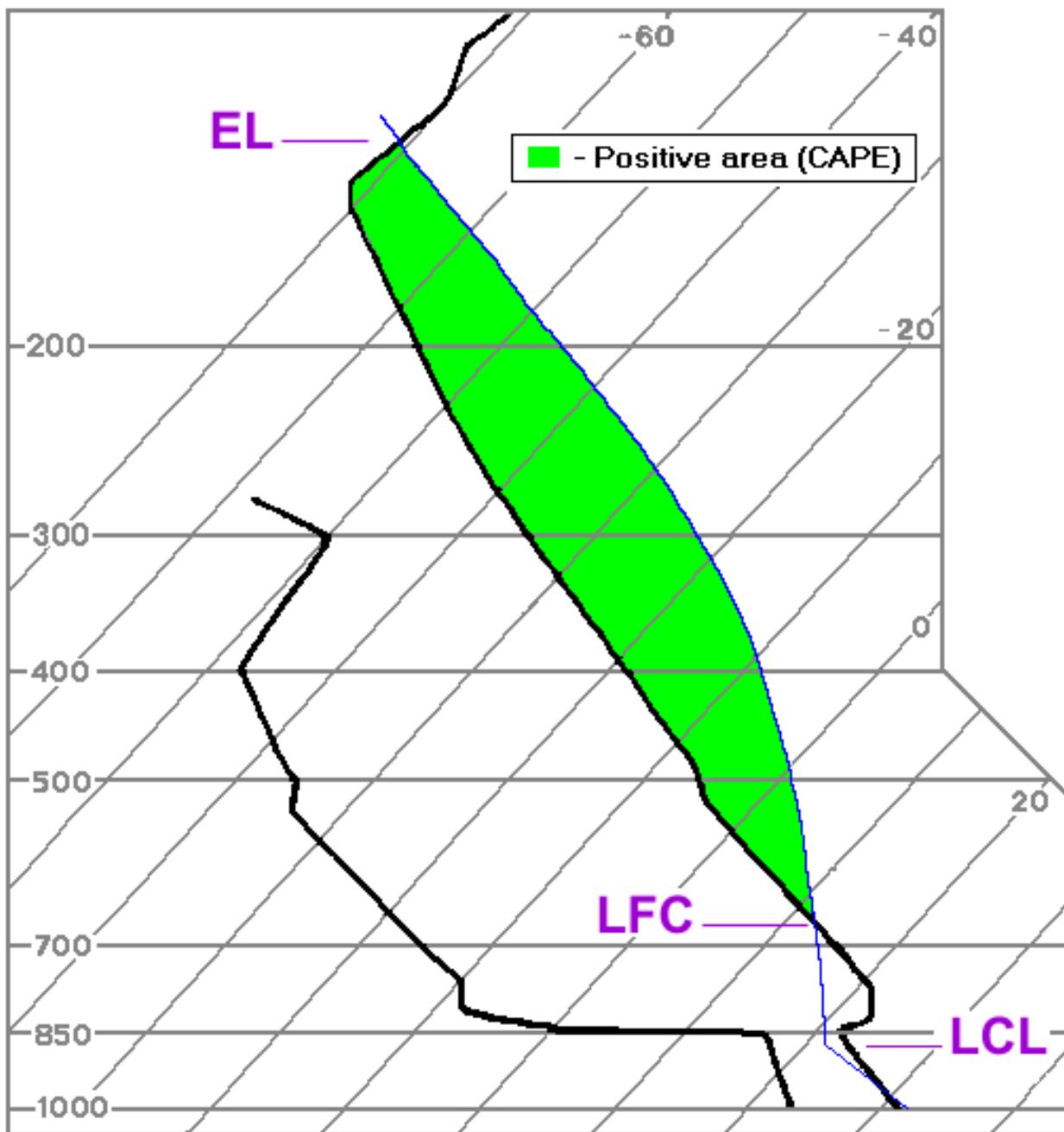
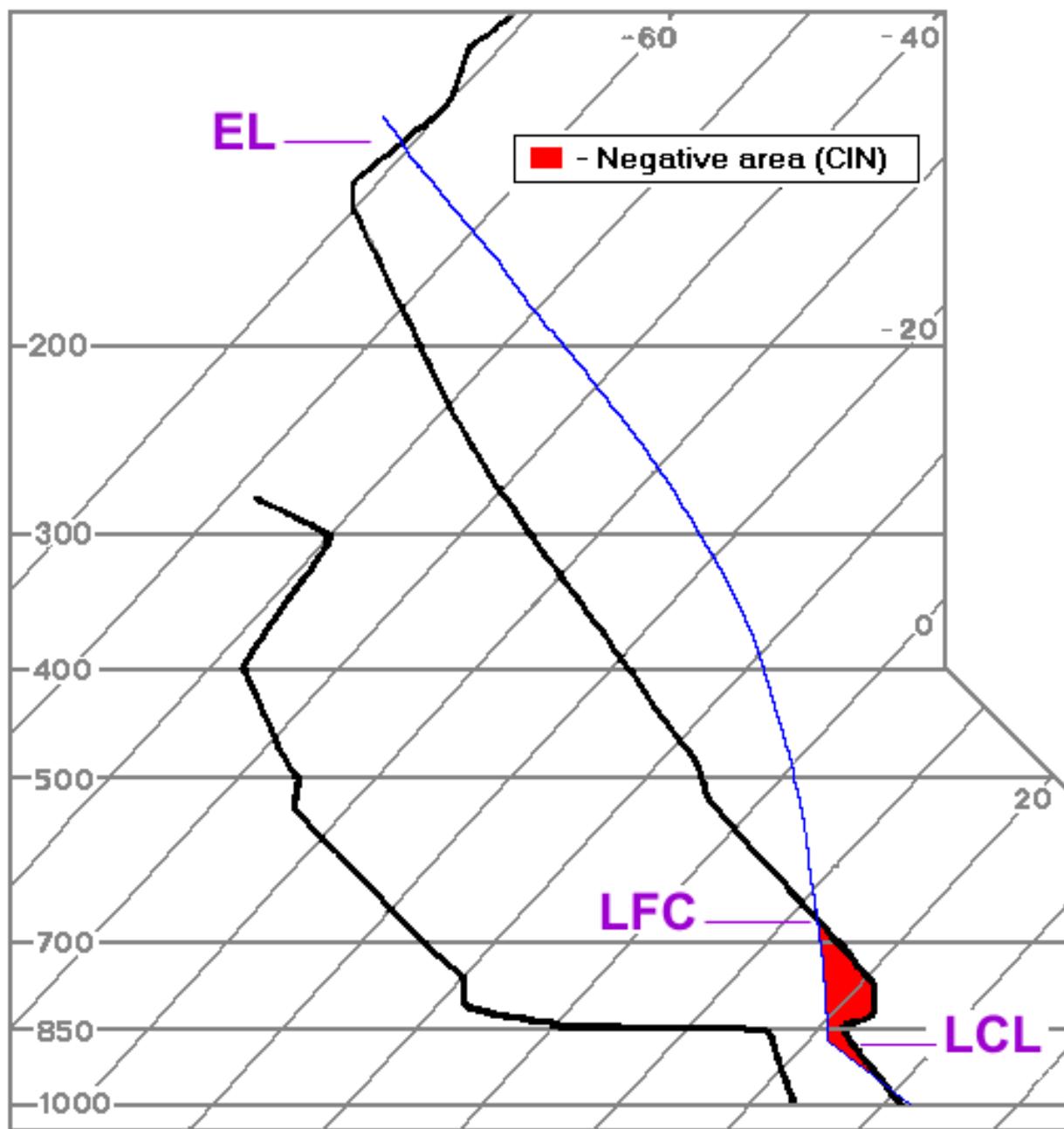


Figure 1: Positive Buoyancy Above Level of Free Convection (CAPE)<sup>27</sup>

Convective inhibition, or CIN, is the opposite of CAPE. In simple terms, CAPE measures instability, while CIN measures stability.<sup>20</sup> CIN measures how much energy is necessary to raise a parcel of air above the level of free convection to a state of positive buoyancy (Fig. 2), which would make the parcel unstable.<sup>21</sup>



**Figure 2: Negative Buoyancy Between Lifting Condensation Level and Level of Free Convection (CIN)<sup>27</sup>**

Different studies have been conducted to investigate how the effects of global warming would affect CAPE values. Using climate models that examine increased moisture, sea temperatures, and surface temperatures scenarios, CAPE values are predicted to increase; one model showed increases in both CAPE and CIN in a warmer climate model.<sup>21</sup> Climate change has impacted and is expected to keep impacting the frequency of extreme weather events.<sup>22</sup> CAPE and CIN are two metrics that can be used to forecast atmospheric instability, and thus, the intensity and probability of extreme weather.<sup>21</sup> By studying trends in CAPE and CIN from radiosonde data, overall trends in extreme weather, one significant impact of climate change, can be observed.

## II. Methods

### A. Database Background

The radiosonde data were extracted from the University of Wyoming's Upper Air Data from the years 1990 to 2023, collected from Station 72520 (PIT) located near Pittsburgh, Pennsylvania. The database consists of twice daily observations, including station information and sounding data. Sounding columns provide radiosonde information on the column of air that the weather balloon travels through. The sounding columns contain atmospheric pressure (hPa), geopotential height (m), which is the altitude with respect to the potential energy between a point and sea level,<sup>28</sup> temperature (°C), relative humidity (%RH), mixing ratio (g/kg), which is the amount of water vapor in the air, wind speed (knots), among other profiles. The data were extracted from the Wyoming Upper Air webpage using a custom Python script that uses html web scraping tools provided within the BeautifulSoup4 library. The extracted web data was placed into a database that was split into two folders, Profiles and Indices, which contained individual files organized by day and time of the observation. The Profiles consisted of the data from the sounding columns described above, and the Indices included station information and other soundings, such as CAPE and CIN and water vapor mixing ratio within the near-surface atmospheric mixing layer. Profiles measure several weather conditions at different heights in the sounding columns, while Indices are calculated using values from the Profiles.

### B. Data Preprocessing

To better visualize the data and observe trends, the dataset was split by Day Number, which assigns a number from 1 to 365 for each day in the years 1990–2023, with Day 1 corresponding to January 1st and Day 365 corresponding to December 31st. The data were further organized if necessary, with it potentially being grouped into seasons or months to examine seasonal trends in the data across the 34 years. The dataset was also grouped into five-year periods, from the first period of 1990–1994 to the most recent period of 2020–2023 to allow for more careful analysis of changes in weather trends over time. Five-year intervals were chosen to mitigate the potential impacts of El Niño. El Niño is a weather phenomenon that occurs every three to five years that results in warmer and drier conditions in Pittsburgh.<sup>29</sup> By using five-year intervals, El Niño would be generally accounted for, as its impacts would affect each period equally.

### C. Initial Data Visualization

Data visualization methods were employed to better understand the database and the impacts of the weather profiles. The dataset was first bounded from 10–500 meters geopotential height to examine surface weather trends. Then, from the bounded dataset, temperature, pressure, relative humidity, and wind speed were graphed according to season, day number, and height. These summary figures were graphed alongside the full range of data to examine how each period deviated from the overall dataset. The dataset was bounded from 9500–16000 meters geopotential height to examine weather trends in the tropopause region. Summary graphs were also produced for the tropopause-bounded dataset following the same format as the surface weather trends.

## D. Analysis Overview

### 1. Anomalies

Anomalies, or departures from averages, are widely used in atmospheric science because of the variability of weather profiles, and anomalies are a good measure of how the data departs from a long-term average. Anomalies generally provide a more accurate description of variability within a profile over a set period than the raw data itself.<sup>4</sup>

The median values for surface temperature, wind speed, relative humidity, and pressure were calculated for the full range of data and each five-year period. Medians were used rather than means to minimize the impacts of drastic outliers. Anomalies were then calculated using the median values.

Boxcar averaging was employed for each variable to remove extra noise in the signal; this technique replaces a group of consecutive points with their average, thus amplifying the signal. Irregularities were smoothed out via boxcar averaging, allowing trends to be more identifiable. With a set box range of 10 days, the boxcar average was calculated for each data point per variable in the entire dataset and each five-year period.

The anomalies were then calculated by subtracting the boxcar average of each point in the overall dataset from the boxcar average of the five-year period. This process was repeated for each five-year period to understand how the observations changed over time, while also understanding the variability within the profile.

### 2. Convective Available Potential Energy (CAPE) and Convective Inhibition (CIN)

After extracting the CAPE values of each observation from the Indices files and compiling them into a file organized by day number, the frequency of CAPE values above specific thresholds was counted to observe the change in convection over the past 34 years. As any CAPE value greater than 0 J/kg indicates instability within the atmosphere<sup>25</sup>, the first threshold examined found values greater than 0 and less than 1000 J/kg. Additionally, thresholds of 1000–2500 J/kg and 2500–3500 J/kg were used, as observations between 1000 and 2500 J/kg indicate moderate instability.<sup>25</sup> A similar process was followed for Convective Inhibition. The thresholds used were -50 to 0 J/kg for weak stability, -200 to -50 J/kg for moderate stability, and less than -200 J/kg for strong stability. These thresholds were chosen based on the distribution of data that was found using a histogram of the CIN values.

### 3. Mean Mixed Layer Mixing Ratio (MLMR)

After extracting the MLMR from the Indices, the data were first grouped by Day Number. The data were averaged for each month of each of the 34 years, with five years of averages for each month, to examine seasonal and monthly trends. This process was repeated for each five-year period from 1990-2023 and the full range of data. The five-year averages were then graphed alongside the average of the full dataset by month. Anomalies were then found using the same method as described earlier.

### 4. Lapse Rates

Using the bounded data for the tropopause from 9500 to 16000 meters, the lapse rate was found using a Python script that placed the data into a dataframe and then calculated the difference between each consecutive data point of the temperature data and the height data. The calculated differences were

appended to new columns in the dataframe. A ratio of the change in temperature,  $dT$ , to the change in height,  $dz$ , was then calculated and appended to a new column in the dataframe,  $dT/dz$ . Lapse rates of the tropopause were then calculated by finding the first value of  $dT/dz$  less than or equal to  $2^{\circ}\text{C}/\text{km}$  and recording the corresponding height. Anomalies were then found using the same method for tropopause height, pressure, temperature, and wind speed to examine trends at a different atmospheric layer.

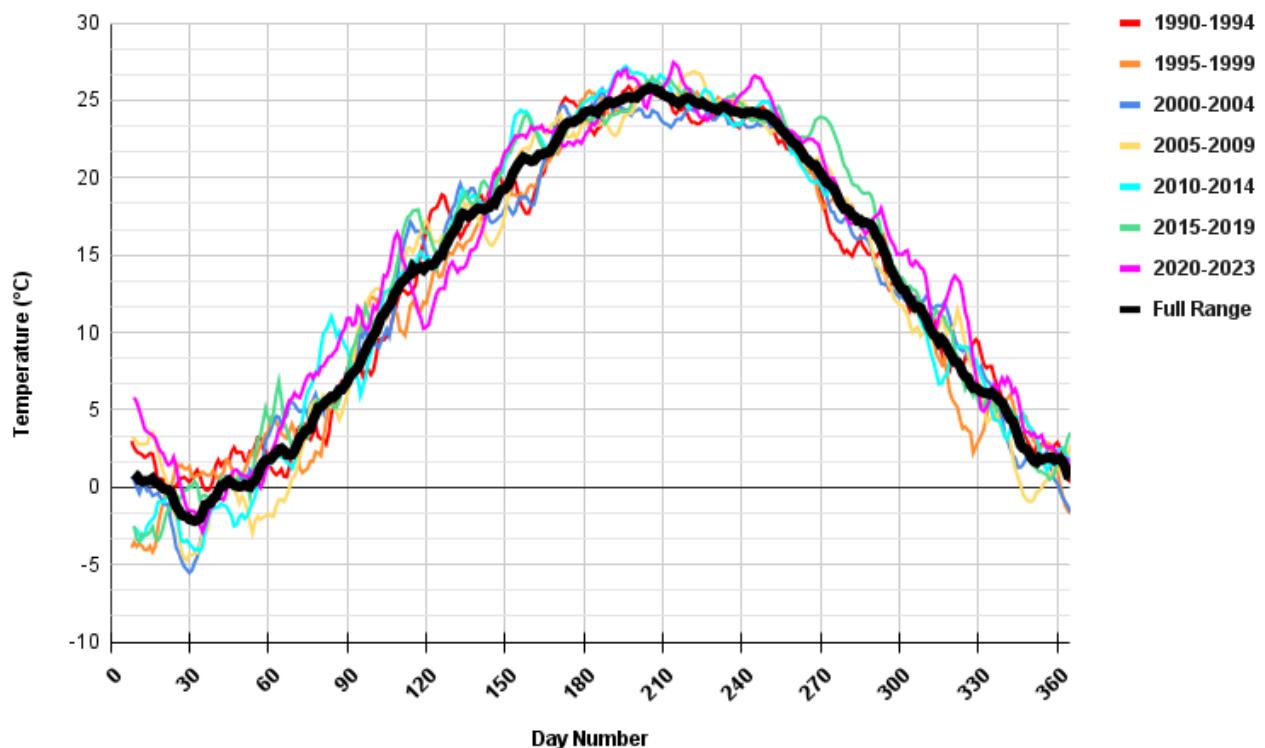
### III. Results & Discussion

#### A. Surface Summary Figures

Summary figures for surface weather trends were produced to determine whether to further investigate the specific meteorological profile. If there appeared to be a trend in the median surface profile summary figure for the span of the entire dataset, anomaly figures were produced to investigate the profile trend more closely.

##### 1. Temperature

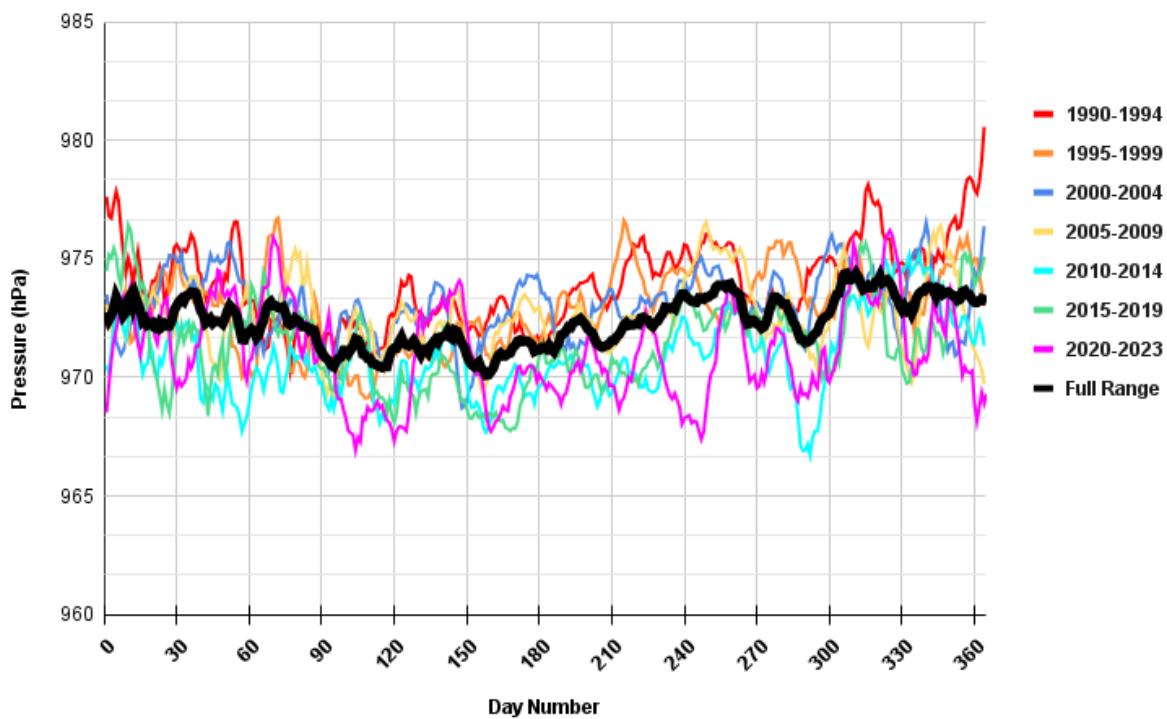
For the 2020–2023 year range, most of the peaks were above the full range dataset, indicating that temperatures were higher for these years than in previous years. Many peaks for the 2010–2014 and 2015–2019 periods were above the full range as well. As shown in Figure 3 below, a slight trend can be seen with increasing temperatures in more recent five-year periods, displaying the global impacts of climate change on temperature to a lesser extent on the local scale.



**Figure 3: Median Surface Temperature 1990-2023**

## 2. Pressure

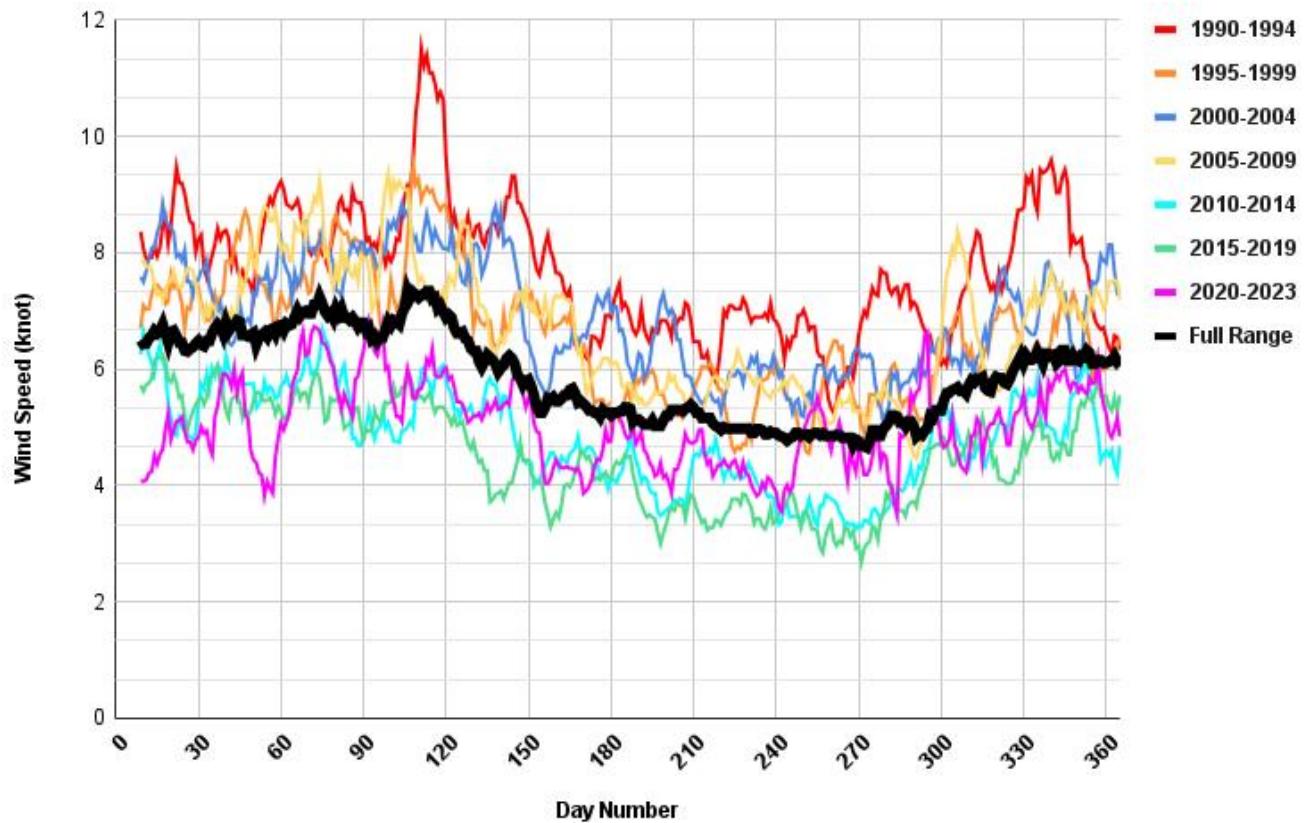
For the past thirteen years, Figure 4 displays a trend of decreasing median surface pressure. From 2010 onwards, surface pressure appears to have been decreasing, with the period from 2020–2023 generally being lower than the previous periods and the full range. The emergence of a trend in the summary figure indicates a need for further analysis of surface pressure.



**Figure 4: Median Surface Pressure 1990-2023**

## 3. Wind Speed

Figure 5, a summary figure of wind speed, displays that wind speeds have been lower, on average, for recent years as opposed to the early years of the dataset. It was found that wind speed was generally increasing from 1990-2009, but from 2010-2023, it seems that wind speeds in Pittsburgh have been declining. The period from 2020-2023 falls below the median surface wind speeds of the full range of data at times, indicating that wind speeds are lower now than they generally have been for the past 34 years.



**Figure 5: Median Surface Wind Speed 1990-2023**

#### 4. Relative Humidity

There was not a clear periodic trend seen with relative humidity, but a seasonal trend was observed in Figure 6. The lower trends in the warmer summer months and the higher trends in the winter months correspond to the influence of temperature on relative humidity. Because warmer air can hold more water vapor than cooler air, relative humidity decreases in the summer months and increases in the winter months. However, clear trends are not apparent.

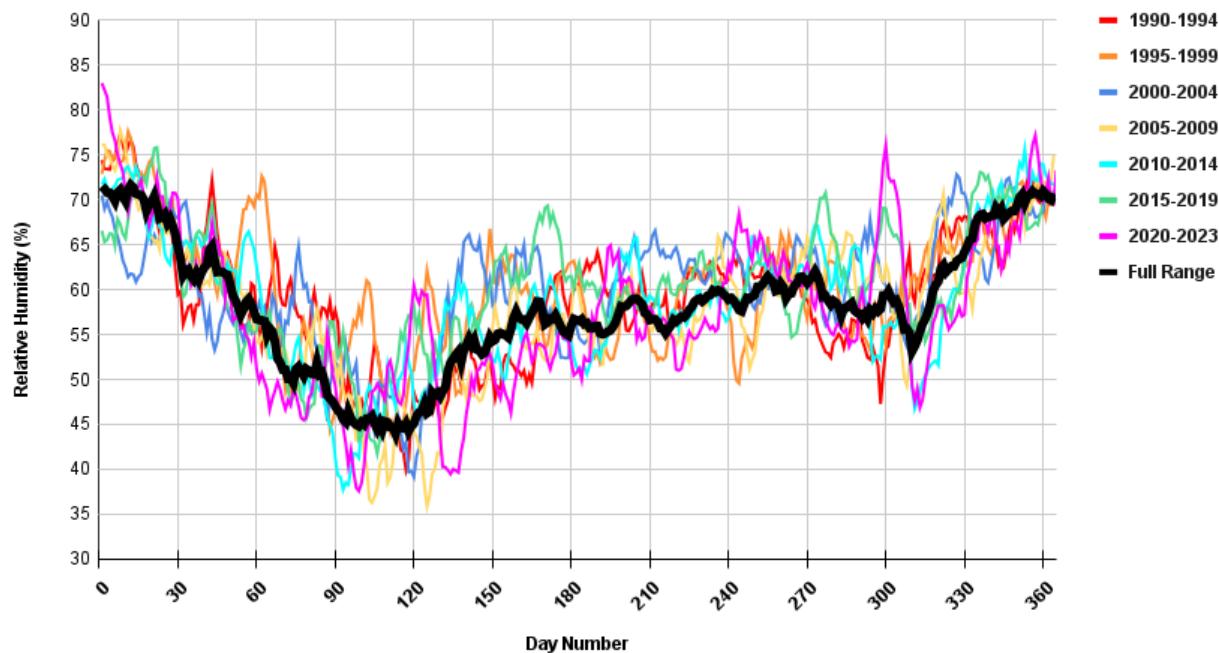


Figure 6: Median Surface Relative Humidity 1990-2023

## B. Surface Anomalies

### 1. Temperature

For nearly all months in the past eight years and the majority of 2010-2014, positive anomalies were observed for temperature, as seen in Figure 7. This indicates that surface temperatures are increasing, which corresponds to the known trend of global warming. However, this profile exhibits great variability in more recent years, due to the sensitive nature of temperature. The recent variability thus results in a weak trend seen between the periods.

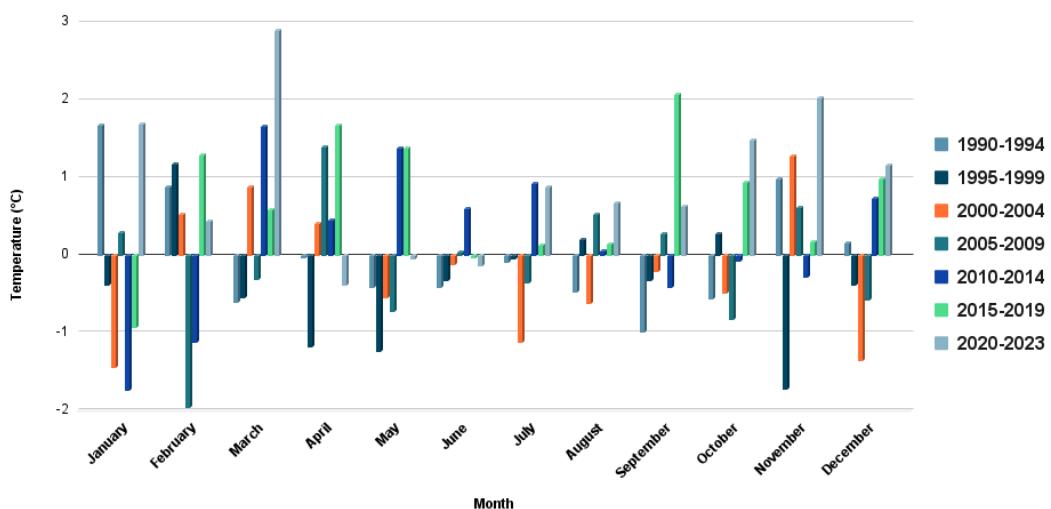
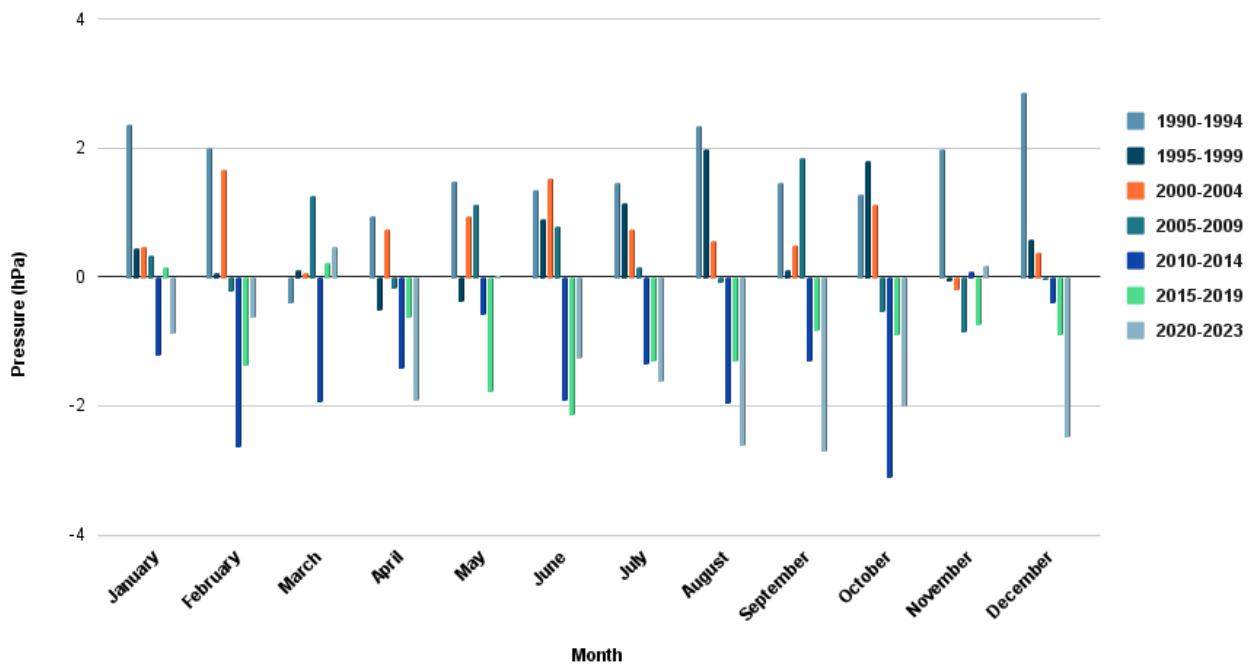


Figure 7: Median Monthly Surface Temperature Anomaly

## 2. Pressure

Surface pressure was found to have generally negative anomalies for the periods between 2010 and 2023 and generally positive anomalies for the periods between 1990 and 2004, with 2005-2009 mainly having anomalies near zero. This indicates a trend of decreasing pressure within the surface layer. While Figure 7 did not show a distinct periodic trend for temperature, Figure 8 provides support for warming surface temperatures locally. Temperature is affected by many confounding variables, thus making its local trends more variable. However, because pressure is not as prone to being affected by confounding variables, pressure trends can be used as an indicator for temperature trends, as pressure is a function of temperature. Decreasing pressure is a result of less dense air, and air density decreases with warmer temperatures. Thus, a decreasing surface pressure trend provides strong support for an increasing surface temperature trend in Pittsburgh.



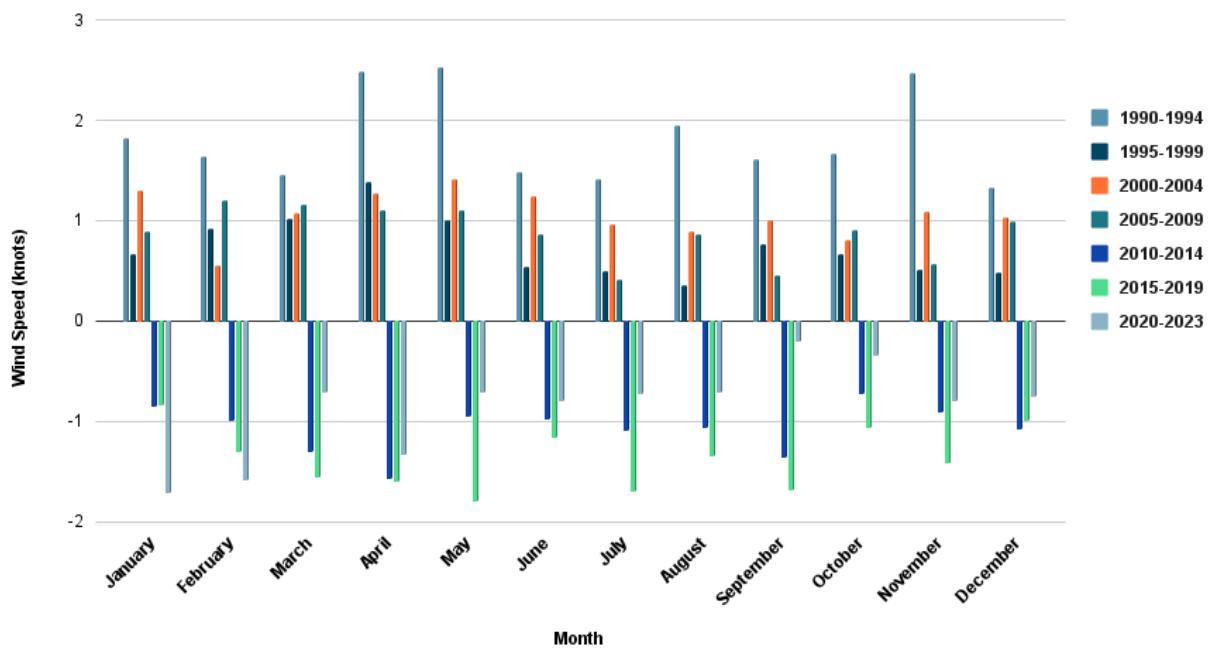
**Figure 8: Median Monthly Surface Pressure Anomaly**

## 3. Wind Speed

Figure 9 displays the median monthly surface wind speed anomalies from 1990 to 2023. A clear distinction is seen between the periods with negative anomalies and those with positive anomalies. The three most recent periods have all negative anomalies, whereas the first four periods contain all positive anomalies. This is indicative of a trend of decreasing wind speed over the past 34 years.

Global Terrestrial Stilling is a global phenomenon where surface wind speeds have been decreasing since the 1980s.<sup>24</sup> It is believed that Terrestrial Stilling is a result of climate change, although the cause is unknown and yet to be proven.<sup>24</sup> Figure 9 provides support for local impacts of Global Terrestrial Stilling, as it shows clear trends of decreasing wind speed. Recently, however, there has been a global trend of a reversal in Global Terrestrial Stilling, with increasing temperatures since around 2010.<sup>24</sup> The wind speed anomaly figure displays this reversal trend as well, with the 2015–2019 period having less negative

anomalies than 2010–2014 and the 2020–2023 period having less negative anomalies than 2015–2019. Wind speeds have been rising once again, supporting the reversal of Global Terrestrial Stilling as well.



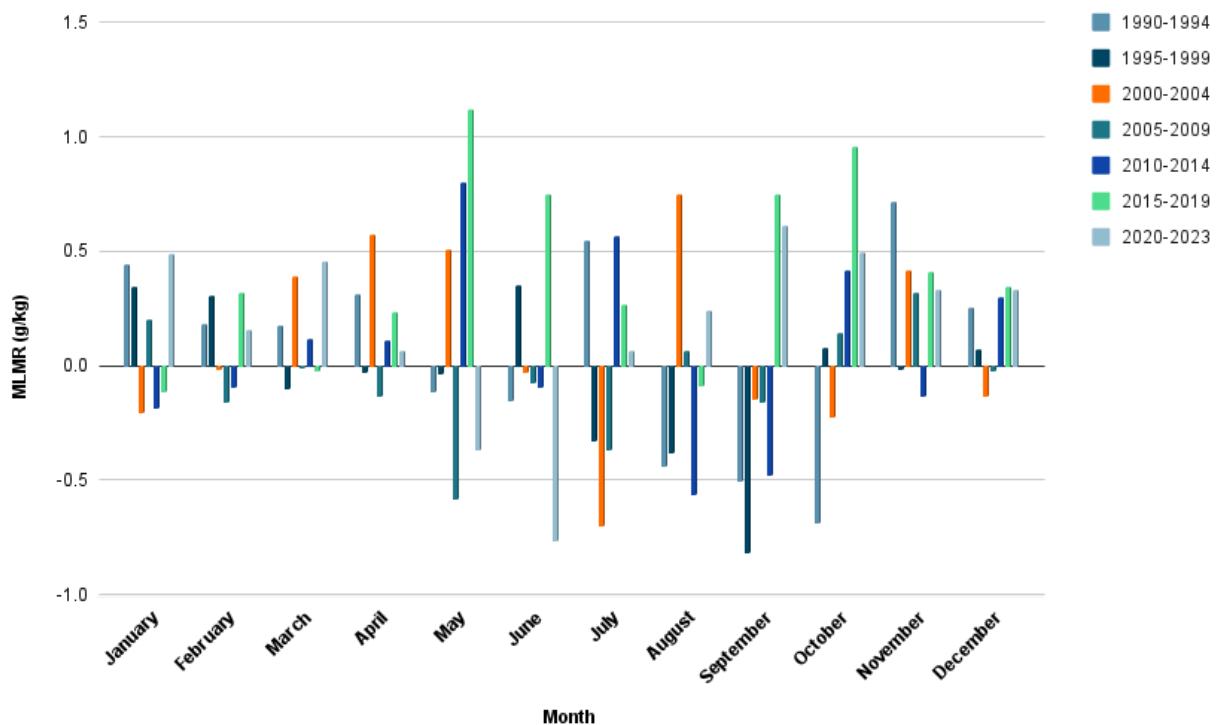
**Figure 9: Median Monthly Surface Wind Speed Anomaly**

#### 4. Relative Humidity

Because Figure 6 showed no clear trends in surface relative humidity, it was determined that an anomaly figure for the profile was not necessary. Relative humidity will not have distinct periodic trends due to it taking into account the maximum amount of water vapor the air can hold based on the corresponding temperature.

#### C. Mixing Layer

The Mean Mixing Layer Mixing Ratio anomalies in Figure 10 did not appear to have a clear periodic trend, but there does seem to be a seasonal trend emerging. While the 2015–2019 period shows large spikes seen corresponding with large positive anomalies in absolute humidity, the other periods do not show the same pattern in the anomalies. Globally, MLMR is predicted to increase, as it is analogous to absolute humidity. Warmer air has the potential to hold more water vapor, so if temperatures are increasing globally, the MLMR should increase as well. However, this global trend is not strongly supported on a local scale, as there appears to be a weak periodic trend for MLMR.

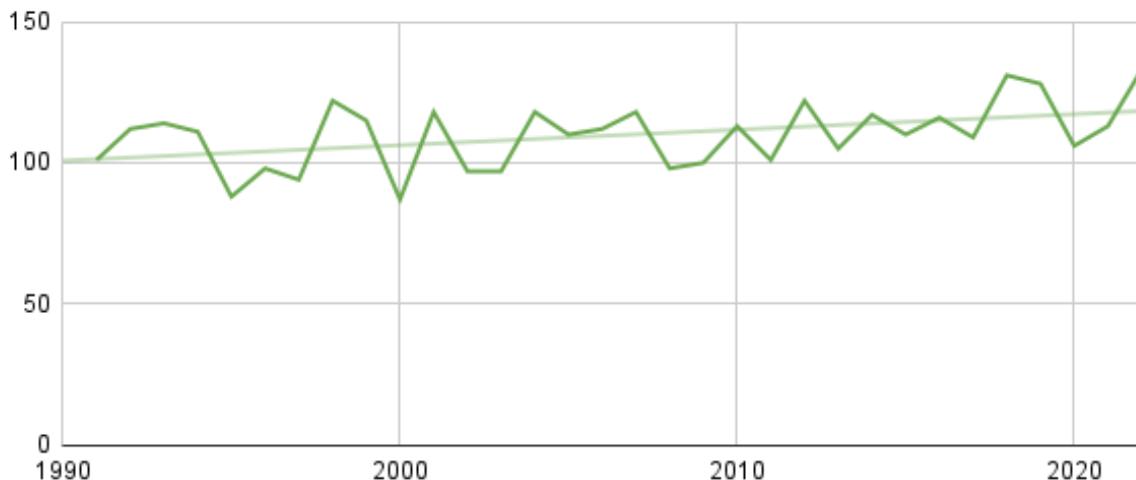
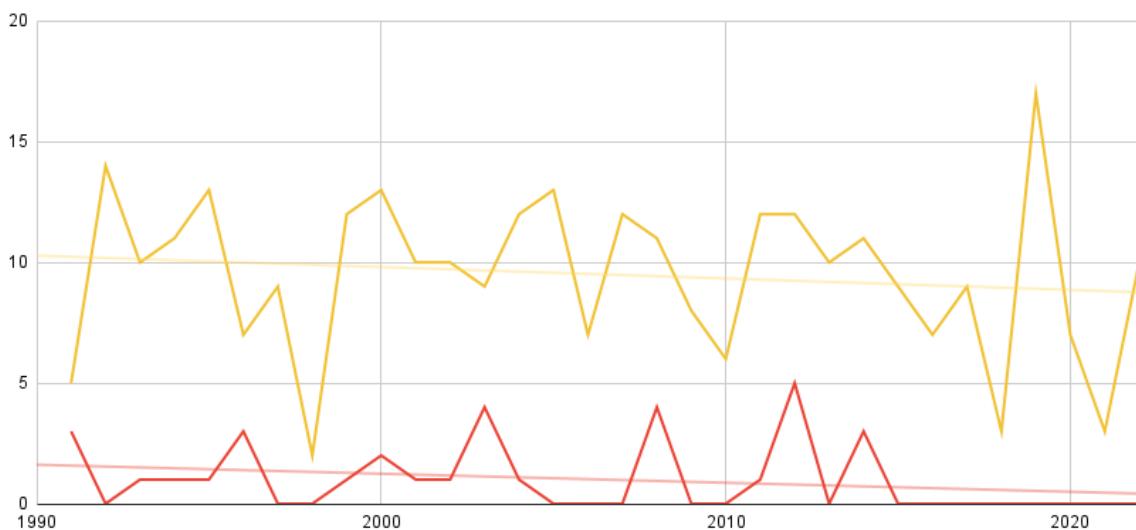


**Figure 10: Median Monthly Surface MLMR Anomaly**

## D. CAPE & CIN

### 1. Convective Available Potential Energy

From Figure 12, the frequency of moderate to strong CAPE values marginally decreased throughout the 34-year dataset, whereas Figure 11 shows the frequency of weak CAPE values marginally increasing. Globally, CAPE is predicted to follow a trend of increasing values, indicating more convective energy in the atmosphere. Although the local trend found does not match up with the predicted global trend for CAPE, it does show an increasing frequency of lower-energy storms and a decreasing frequency of higher-energy storms.

**Figure 11: CAPE Weak Instability 1990-2023****Figure 12: CAPE Moderate and Strong Instability 1990-2023**

## 2. Convective Inhibition

The CIN graphs, Figures 13 and 14, showed some correlation to the results found in the CAPE graphs. The frequency of days with less negative CIN values, indicating weak stability, was found to slightly decrease but overall stay fairly constant (Fig. 13). This corresponds to the relatively constant frequency of days with high CAPE values, which indicated strong instability. The frequency of days with more negative CIN values (Fig. 14), indicating strong stability, had a slight trend of increasing frequency. This corresponds to the results of Figure 11, which showed a slight trend of increasing levels of low convective energy. Overall, the trends in CAPE and CIN showed that there are more days with present but weak convective energy.

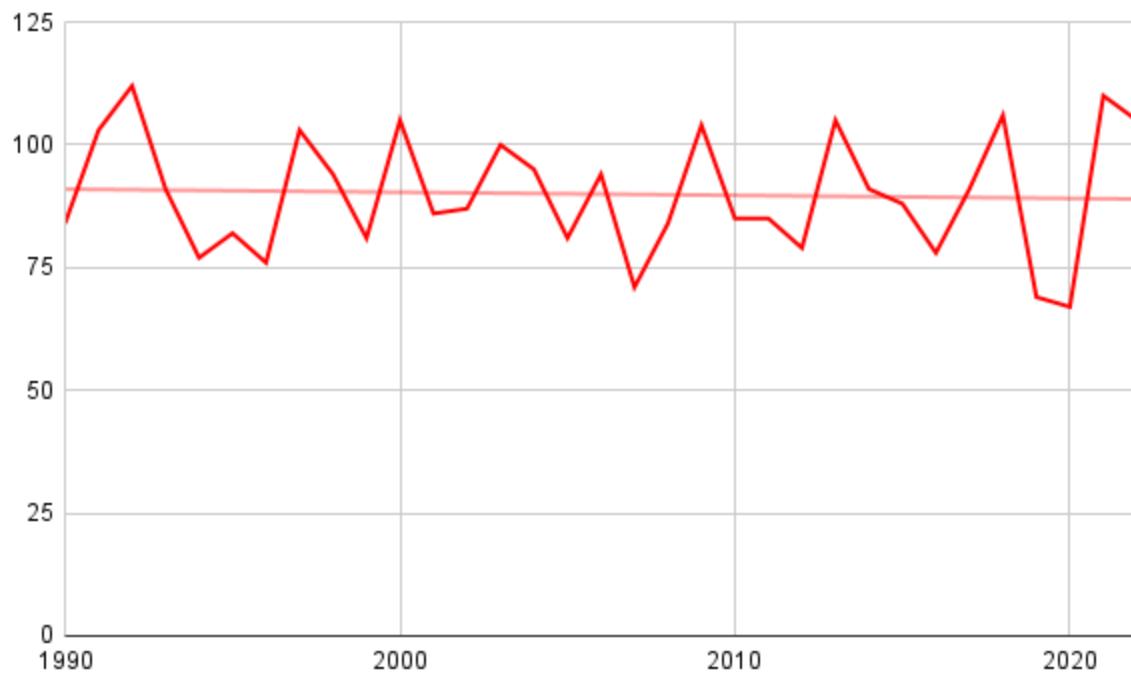


Figure 13: CIN Weak Stability 1990-2023

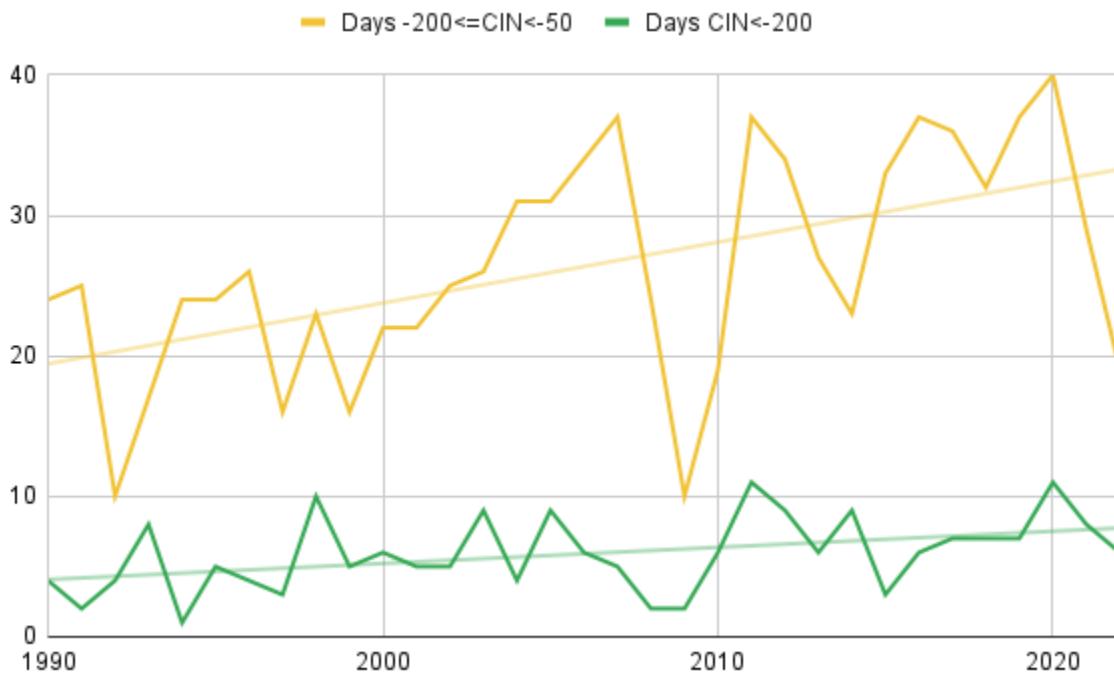
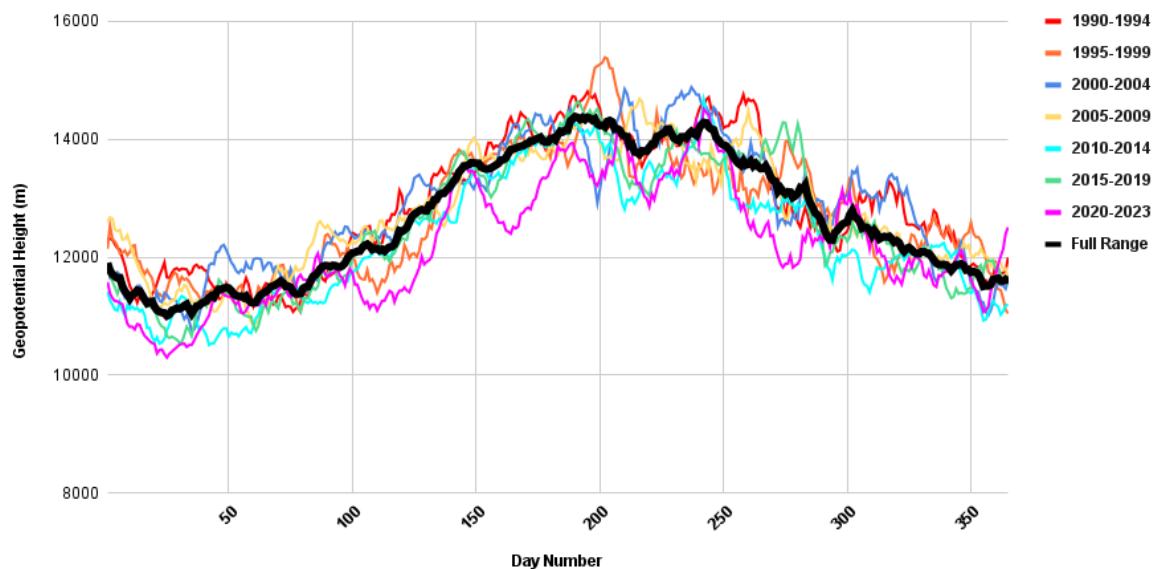


Figure 14: CIN Moderate to Strong Stability 1990-2023

## E. Tropopause Summary Figures

### 1. Height

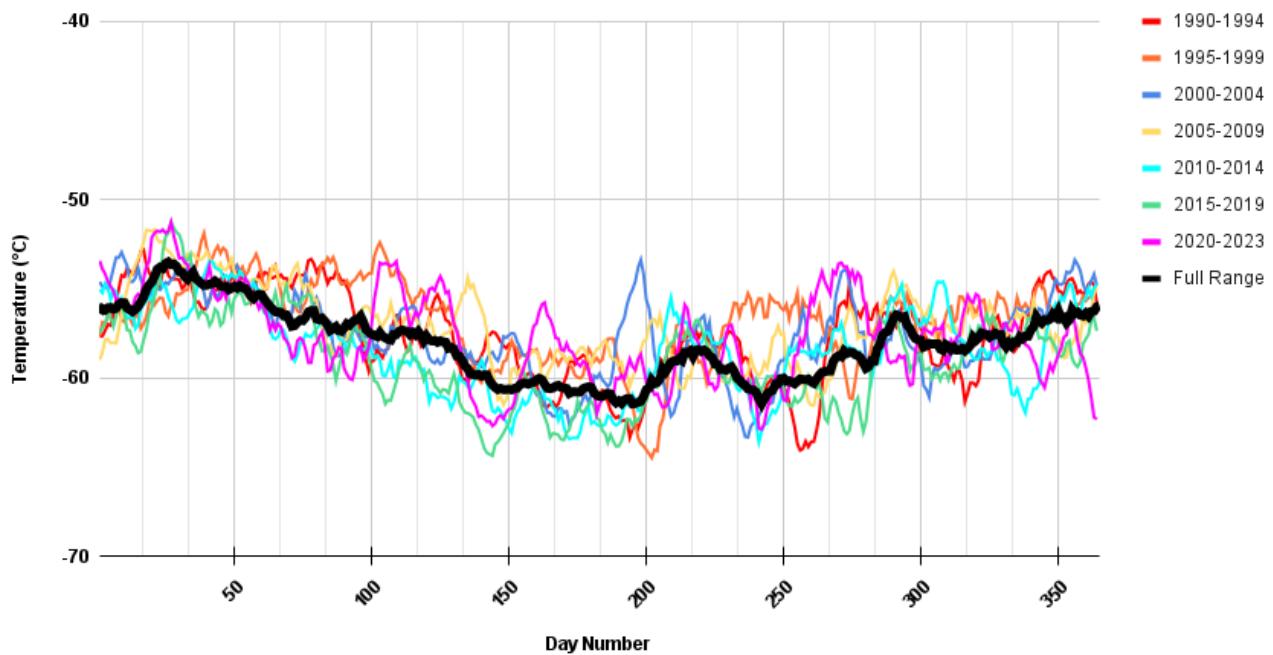
From Figure 15, tropopause height over the past 34 years overall follows a similar seasonal trend of increasing during the summer months when the temperature is generally warmer and decreasing in the winter months when the temperature is generally cooler. The lines for the three most recent periods in Figure 15 are seen to be generally lower than the line for the full range of data, indicating that tropopause height may be decreasing locally.



**Figure 15: Median Tropopause Height Summary 1990-2023**

### 2. Temperature

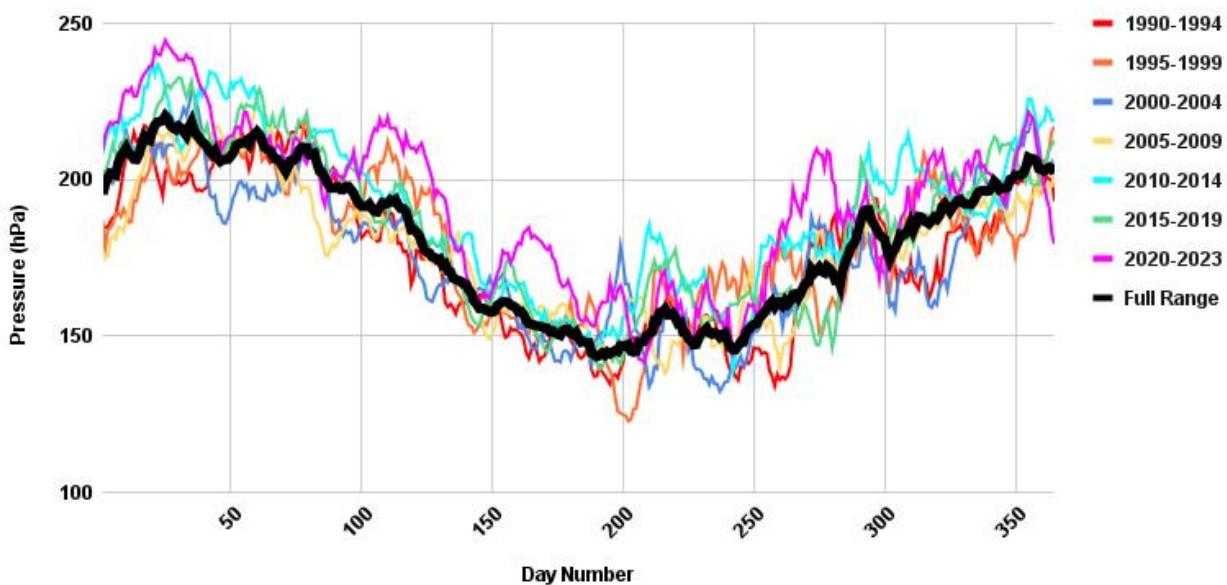
There appears to be increased variability in tropopause height in the more recent periods, making it difficult to discern clear trends. There does not appear to be a strong seasonal trend for tropopause temperature. A clear trend for temperature cannot be found through the summary figure; however, due to the variable nature of temperature, there is still a need for further examination of the tropopause temperature profile despite the lack of support from Figure 16.



**Figure 16: Median Tropopause Temperature Summary 1990-2023**

### 3. Pressure

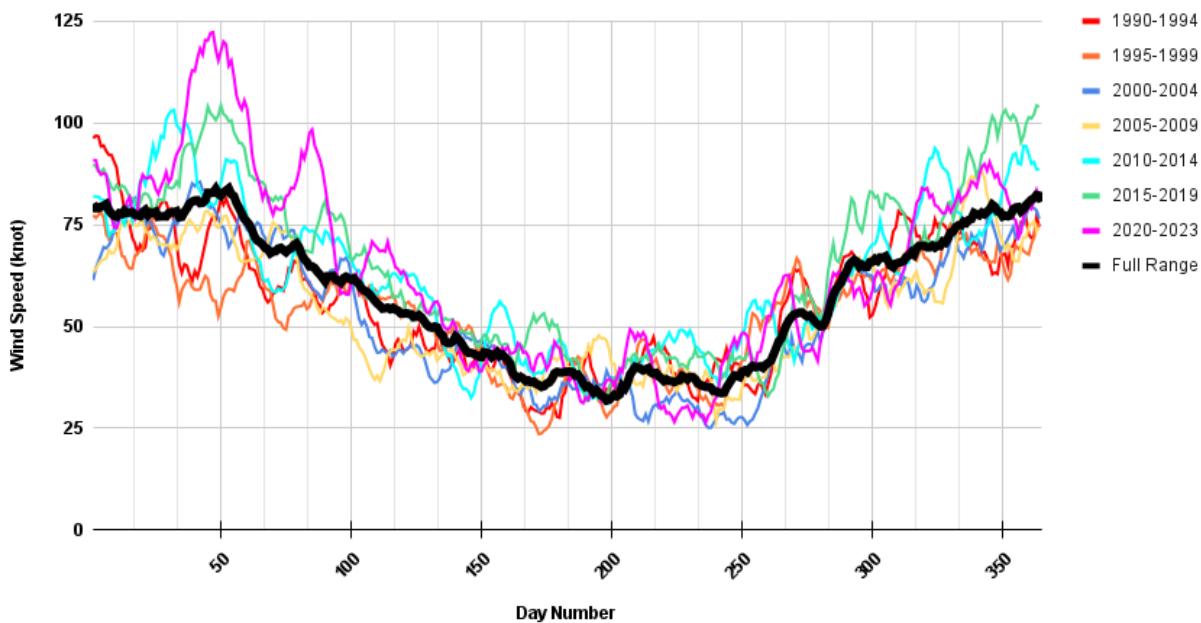
The tropopause initially displays a pattern of increasing pressure as the time periods become more recent. The more recent periods grow higher than the line for the full range in Figure 17, with 2020–2023 generally having higher pressure than the earlier time periods and the overall dataset.



**Figure 17: Median Tropopause Pressure Summary 1990-2023**

#### 4. Wind Speed

The tropopause appears to have a pattern of decreasing wind speeds from Figure 18. Wind speeds in the most recent period look to be higher than the previous periods. The wind speeds of the first four periods generally are decreasing, but they then appear to increase again through the last three periods.



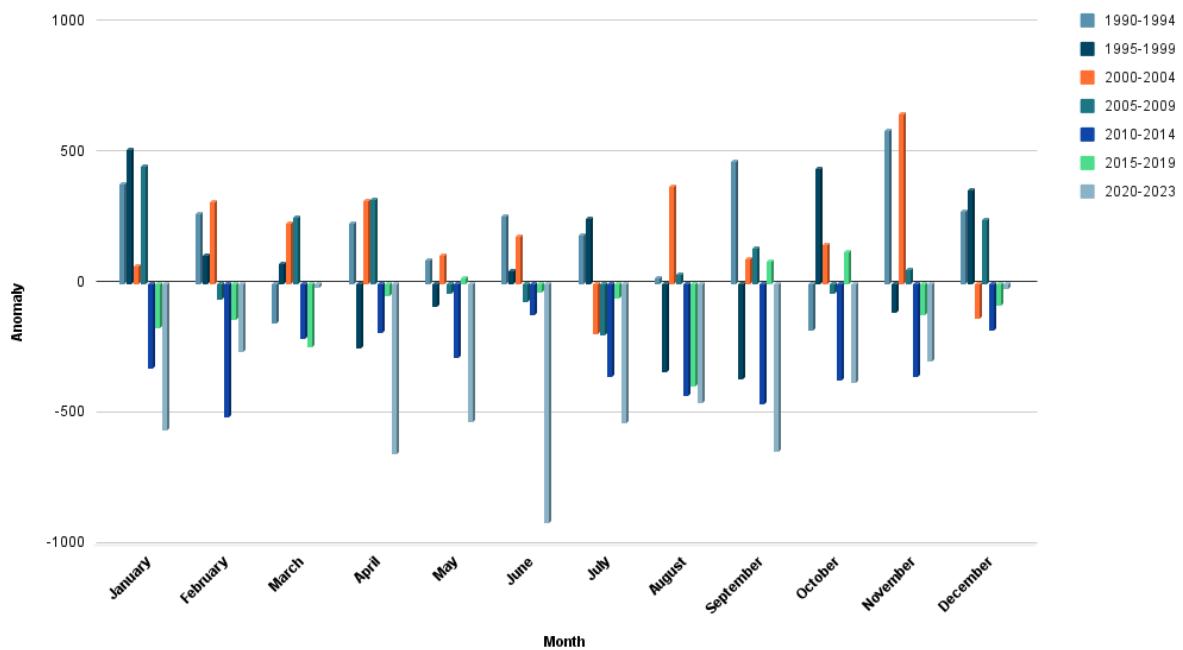
**Figure 18: Median Tropopause Wind Speed Summary 1990-2023**

### F. Tropopause Anomalies

#### 1. Height

Tropopause height is a strong indicator globally for climate change, as tropospheric warming and stratospheric cooling should cause the tropopause height to increase. However, this trend is not seen locally. Rather, a trend of decreasing tropopause height was found in Figure 19. The three most recent periods appear to have mainly negative anomalies, indicating that tropopause height is decreasing in Pittsburgh.

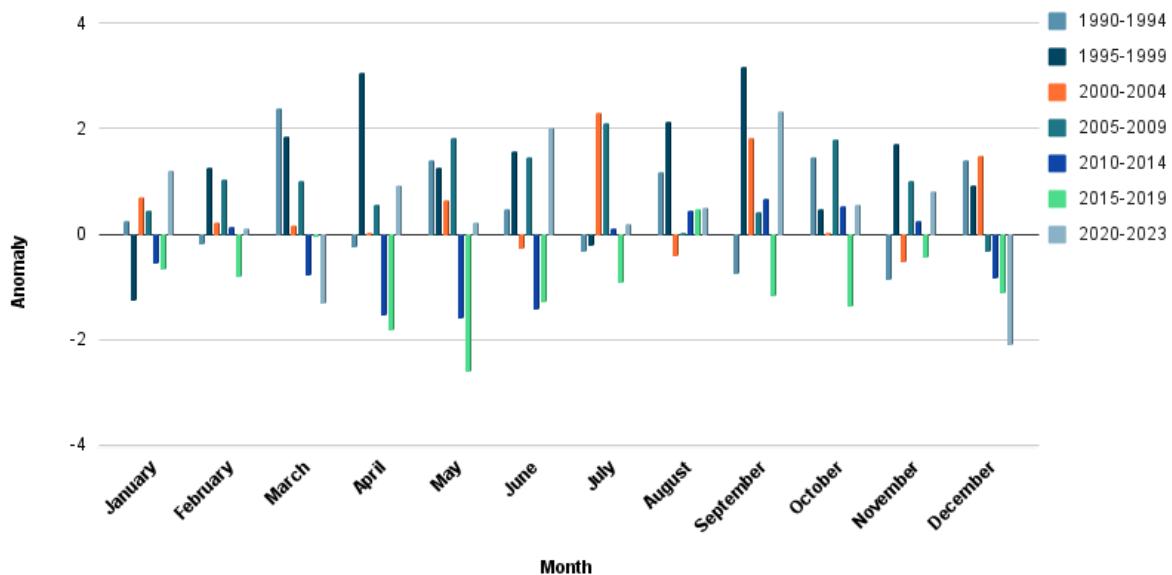
Due to the sensitivity levels of the radiosonde sensors used, gaps are present in the readings that ultimately result in inaccuracies in the calculations for lapse rate. Thus, appropriate height measurements are not recorded either from the lapse rate calculations, giving rise to a trend of decreasing tropopause height rather than following the global trend of increasing height.



**Figure 19: Median Monthly Surface Height Anomaly**

## 2. Temperature

Figure 20 reveals more positive anomalies in the earlier periods for tropopause temperature as compared to the greater proportion of negative anomalies in the recent periods. Specifically seen in the 2010-2014 and 2015-2019 periods, a trend of decreasing tropopause temperature emerges throughout the 34-year dataset. There is a lot of noise and variability in the data, however, even after smoothing, which lessens the strength of the trend, particularly in more recent years.

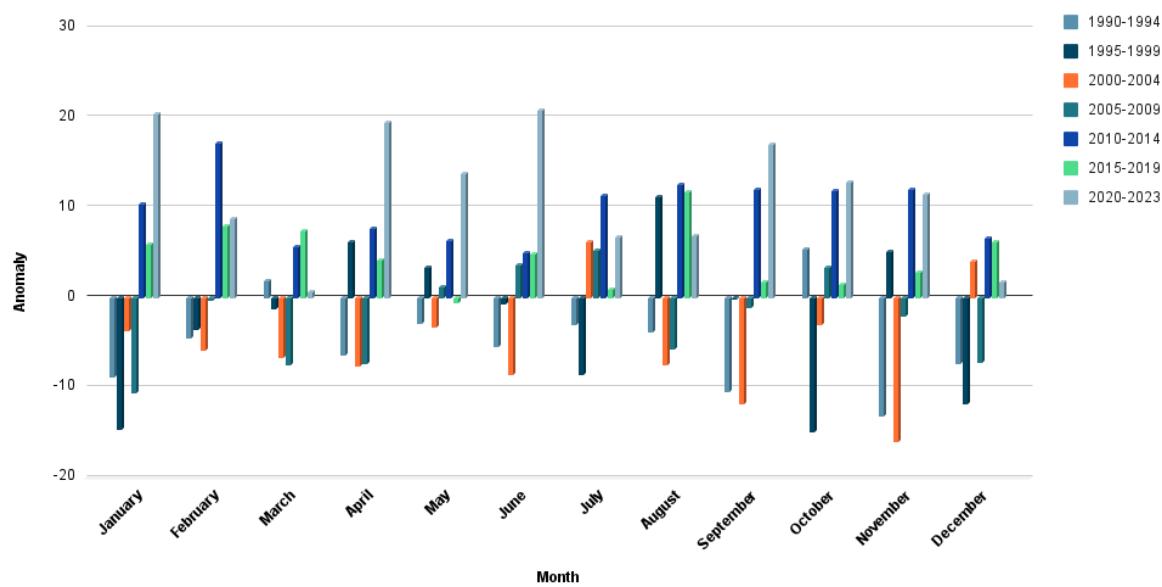


**Figure 20: Median Monthly Surface Temperature Anomaly**

### 3. Pressure

Similarly to the surface layer pressure and temperature trends, the tropopause pressure anomaly figure can be used to examine a trend in tropopause temperature. Figure 21 displays a recent increase in tropopause pressure, as there appears to be a greater proportion of positive anomalies for the four most recent periods as compared to the proportion of positive anomalies for the three earliest periods. Thus, a trend of increasing tropopause pressure can be examined to understand a trend in tropopause temperature.

Because pressure is a function of temperature, increasing tropopause pressure is indicative of decreasing tropopause temperature, providing stronger support for the discussion for Figure 20. A cooler troposphere would result in denser tropospheric air, which leads to increased pressure as seen in Figure 21.

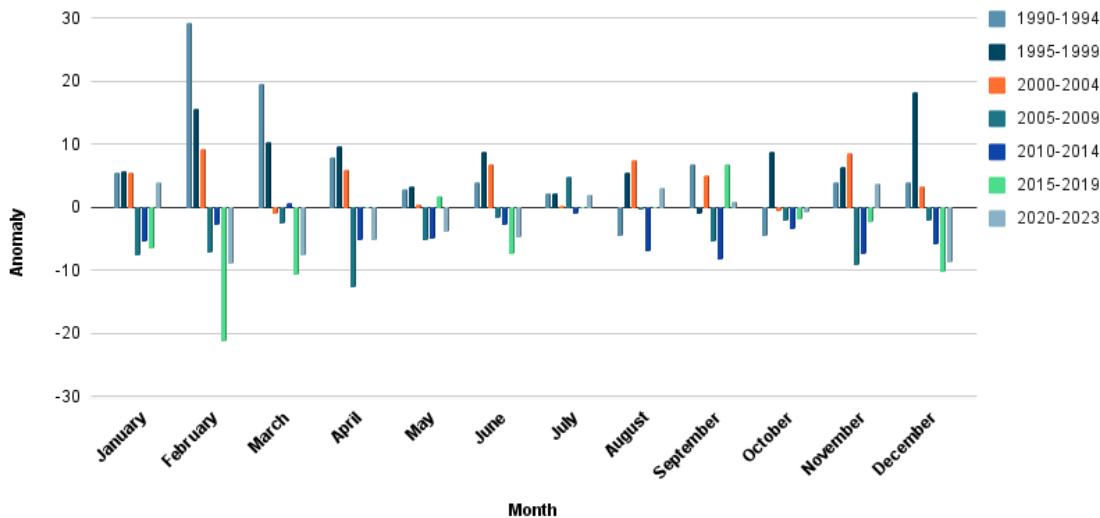


**Figure 21: Median Monthly Surface Pressure Anomaly**

### 4. Wind Speed

The wind speed anomalies appear to be generally negative for the three most recent periods, whereas the anomalies are generally positive for the first four periods in the dataset, as seen in Figure 22. Like the surface layer wind speed trend, wind speeds look to be decreasing throughout the dataset. The wind speeds seem to increase again in the 2020-2023 period, similarly to the surface trend once again. This perhaps could be a result of this period not containing a full five years of data, however.

The similarities in trend between the surface and tropopause wind speed support a theory that a similar mechanism to Global Terrestrial Stilling is occurring in the tropopause. Even the reversal in the trend is seen to some extent with the tropopause wind speed anomalies.



**Figure 22: Median Monthly Surface Wind Speed Anomaly**

## IV. Conclusion

Although several results were ultimately inconclusive, some local impacts followed global trends or resulted in surprising trends. Positive anomalies for median surface temperature in recent years corresponded to the global trend of warming surface temperatures. Negative anomalies in surface pressure also corresponded to warming surface temperatures. For the surface wind speed anomalies, the past thirteen years showed negative anomalies, a trend that could be attributed to Global Terrestrial Stilling. Similarly, in the tropopause wind speed anomalies, the high proportion of negative anomalies for the past thirteen years indicates that a similar mechanism to Global Terrestrial Stilling could be impacting the tropopause. Despite the lack of periodic trends in several profiles, some seasonal trends were observed in these graphs. In Figure 10, the anomaly graph for MLMR, an increase in positive and negative anomalies was observed. Positive anomalies in tropopause pressure indicated that pressure was increasing, leading to the belief that tropopause temperature has been decreasing.

### A. Implications

Comparing the results to global trends, profiles that were extremely variable and easily affected by confounding sources, such as temperature, showed less conclusive results than less variable profiles like pressure, which are affected by fewer factors. However, local impacts of climate change were able to be observed, providing strong support for climate change's overarching impacts. Thus, the trends and conclusions from the study could be used to aid in supporting climate change policy as the local impacts of this global phenomenon were observed.

### B. Limitations

The data were averaged and smoothed in order to compile the data to get conclusive results. Medians and smoothing methods were used to attempt to weaken the impacts of drastic outliers on periodic or seasonal trends, as outliers could skew the dataset in a way that would hide trends. However, this created the limitation of not being able to see how the outer edges of the data, such as outliers, increased over time.

Additionally, because the last period from 2020-2023 did not contain a full five years of data as the other periods did, a skew had the potential to emerge. This could have affected the way trends carried through the last period, as weather patterns could change had there been a full five-year period used.

Furthermore, atmospheric data in general is extremely variable, as the atmosphere is not a controlled space. Unlike a laboratory setting, weather profiles can not be controlled and managed carefully, so confounding variables could affect the trends in certain profiles, particularly temperature. Temperature is influenced heavily by outside factors, which is why trends in pressure had to be used to explain temperature trends.

### C. Further Research

Several additional areas of interest were not yet examined in this analysis of sounding data. Examination of outliers would be useful in understanding the more extreme impacts of climate change on a local scale. Perhaps a closer examination of the outer edges of the data would reveal additional local trends that match with current global trends and predictions.

Additionally, investigating a larger time frame for the database could potentially reveal stronger trends in the weather profiles. Also, using a wider geographical range might provide the potential for the emergence of trends in the database as well.

Furthermore, it would be interesting to examine how El Niño and other similar weather phenomena affect weather trends. While the use of five-year periods did attempt to take into account El Niño, its confounding impacts were not completely eradicated as the phenomena do not occur in a linear timeframe and could still have impacted the data within the five-year period.

Incorporating atmospheric models could help fill in gaps within the database, particularly due to the sensitivity level of the radiosonde sensors used. Because the radiosonde used to measure these profiles did not take readings at every geopotential height, calculations for tropopause lapse rate did not always result in a lapse rate of  $2^{\circ}\text{C}/\text{km}$ , potentially affecting the trend seen for tropopause height. Thus, using an atmospheric model could fill in for the radiosonde profiles corresponding to heights not measured, allowing for a more accurate analysis of trends.

## V. References

1. NASA. (2019, October 16). Earth's atmosphere: A multi-layered cake – climate change: Vital signs of the planet. NASA. <https://climate.nasa.gov/news/2919/earths-atmosphere-a-multi-layered-cake/>
2. Seinfeld, J. H., & Pandis, S. N. (2016). Atmospheric chemistry and physics: from air pollution to climate change. John Wiley & Sons.
3. UCAR Center for Science Education. (n.d.). *The Mesosphere*. UCAR Center for Science Education. <https://scied.ucar.edu/learning-zone/atmosphere/mesosphere>
4. The role of anomalies in climate science. (n.d.).  
<https://sites.stat.washington.edu/peter/statclim/Case%202.html>
5. NASA. (2023, July 26). FAQ: *What is the greenhouse effect?*. NASA.  
<https://climate.nasa.gov/faq/19/what-is-the-greenhouse-effect/>
6. EIA. (2023, July 27). *U.S. Energy Information Administration - EIA - independent statistics and analysis*. Greenhouse gases - U.S. Energy Information Administration (EIA).  
<https://www.eia.gov/energyexplained/energy-and-the-environment/greenhouse-gases.php>
7. US Department of Commerce, N. (2005, October 1). *ESRL Global Monitoring Laboratory - Education and outreach*. GML. [https://gml.noaa.gov/outreach/carbon\\_toolkit/](https://gml.noaa.gov/outreach/carbon_toolkit/)
8. Buis, A. (2022, August 22). *Steamy relationships: How atmospheric water vapor amplifies Earth's greenhouse effect – climate change: Vital signs of the planet*. NASA.  
<https://climate.nasa.gov/explore/ask-nasa-climate/3143/steamy-relationships-how-atmospheric-water-vapor-amplifies-earths-greenhouse-effect/>.
9. NOAA. (2023, June 6). *The earth-atmosphere energy balance | national oceanic and atmospheric ...* NOAA. <https://www.noaa.gov/jetstream/atmosphere/energy>
10. National Oceanic and Atmospheric Administration. (2023a, April 28). *Radiosondes*. National Oceanic and Atmospheric Administration. <https://www.noaa.gov/jetstream/upperair/radiosondes>
11. National Oceanic and Atmospheric Administration. (2023, June 13). *Z-time (coordinated universal time)*. National Oceanic and Atmospheric Administration. <https://www.noaa.gov/jetstream/time>
12. NOAA's National Weather Service. (2023, February 15). *Collecting meteorological data by radiosonde or Weather Balloon*. National Weather Service.  
<https://www.weather.gov/rah/virtualtourballoon>
13. Meng, L., Liu, J., Tarasick, D. W., Randel, W. J., Steiner, A. K., Wilhelmsen, H., ... & Haimberger, L. (2021). Continuous rise of the tropopause in the Northern Hemisphere over 1980–2020. *Science Advances*, 7(45), eabi8065.
14. Tran, L. (2021, June 30). *NASA satellites see Upper Atmosphere Cooling and contracting due to Climate Change*. NASA. <https://www.nasa.gov/feature/goddard/2021/nasa-satellites-see-upper-atmosphere-cooling-contracting-climate-change/>
15. Pisot, P., Sacha, P., Polvani, L. M., Añel, J. A., De La Torre, L., Eichinger, R., ... & Rieder, H. E. (2021). Stratospheric contraction caused by increasing greenhouse gases. *Environmental Research Letters*, 16(6), 064038.
16. US Department of Commerce, N. (2015, June 13). *A Discussion of Water Vapor, Humidity, and Dewpoint, and Relationship to Precipitation*. National Weather Service.  
<https://www.weather.gov/lmk/humidity>
17. *Mixing ratio*. Mixing ratio - Glossary of Meteorology. (2012, January 26).  
[https://glossary.ametsoc.org/wiki/Mixing\\_ratio](https://glossary.ametsoc.org/wiki/Mixing_ratio)
18. *Sounding Station Parameters and Indices*. Sounding station parameters and indices. (n.d.).  
<https://weather.uwyo.edu/upperair/indices.html>
19. Solanki, R., Macatangay, R., Sakulsupich, V., Sonkaew, T., & Mahapatra, P. S. (2019, November

- 5). *Mixing layer height retrievals from MINIMPL measurements in the Chiang Mai Valley: Implications for particulate matter pollution.* Frontiers. <https://www.frontiersin.org/articles/10.3389/feart.2019.00308/full>
20. US Department of Commerce, N. (2015, May 20). *Severe weather topics.* National Weather Service. <https://www.weather.gov/ilx/swop-severetopics-CAPE>
21. Chen, J., Dai, A., Zhang, Y., & Rasmussen, K. L. (2020, February 1). *Changes in convective available potential energy and convective inhibition under global warming.* AMETSO. <https://journals.ametsoc.org/view/journals/clim/33/6/jcli-d-19-0461.1.xml>
22. Martinkova, M., & Kysely, J. (2020, July 25). *Overview of observed clausius-clapeyron scaling of extreme precipitation in Midlatitudes.* MDPI. <https://www.mdpi.com/2073-4433/11/8/786>
23. US Department of Commerce, N. (2015b, June 13). *Discussion on humidity.* National Weather Service. <https://www.weather.gov/lmk/humidity>
24. Zeng, Z., Ziegler, A.D., Searchinger, T. et al. A reversal in global terrestrial stilling and its implications for wind energy production. *Nat. Clim. Chang.* 9, 979–985 (2019). <https://doi.org/10.1038/s41558-019-0622-6>
25. NOAA's National Weather Service. (2015, June 13). *ENV parameters and indices.* National Weather Service. <https://www.weather.gov/lmk/indices>
26. Santer, B. D., Wehner, M. F., Wigley, T. M. L., Sausen, R., Meehl, G. A., Taylor, K. E., ... & Bruggemann, W. (2003). Contributions of anthropogenic and natural forcing to recent tropopause height changes. *science*, 301(5632), 479-483.
27. Penn State Meteorology. (n.d.). *Assessing strong updrafts.* Penn State Meteorology. <https://learningweather.psu.edu/node/87>
28. Oxford University Press. (2023). *Geopotential height.* Oxford Reference. <https://www.oxfordreference.com/display/10.1093/acref/9780199541447.001.0001/acref-9780199541447-e-2577>
29. National Oceanic and Atmospheric Administration. (2009, March 26). *What are El Niño and La Niña?* National Ocean Service. <https://oceanservice.noaa.gov/facts/ninonina.html>

## VI. Acknowledgements

We would like to express our most heartfelt thanks and gratitude to Dr. Barry Luokkala, Dr. Doug Goetz, Ashley Driesbach, Julia Wang, Lora Kallenberg, Ms. Melissa Lessure, and the PGSS Campaign, Inc. for their tremendous help and support throughout the course of this project. We would also like to extend special thanks to our parents, guardians, and other family members for the support they have given us during our time at PGSS. We could not have done this without all of your support.



# A Breath of Fresh Air? Analysis of Outdoor and Indoor Air Quality and Exposure in Pittsburgh

John Beeson, Isabelle Ehrensberger, Carlos Peterson, Mackenzie Vasbinder

## Abstract

Through weather apps and websites typically produced by regulatory agencies, the general public can monitor the air quality in their area. However, the sensors from which this information is derived cover such large radii that they cannot account for local events that impact the air quality in specific towns or neighborhoods. A possible solution to this problem is using low-cost air sensors, which cost substantially less than those used by regulatory monitoring stations and pick up on local events that may impact the air quality. The purpose of this project is to investigate whether these low-cost sensors could be used to accurately provide specific information about the quality of air students were exposed to during their time spent at the Pennsylvania Governor's School for the Sciences held on the Carnegie Mellon University campus from July 2nd to August 5th in 2023. Through the use of RAMP sensors and analysis of the data collected, it was determined that, on average, the concentration of PM<sub>2.5</sub> was higher outside than inside, and the concentration of CO<sub>2</sub> was higher inside than outside. Such analysis also concluded that several events, such as yard work and the spread of Canadian wildfire smoke, also heavily impacted the air quality on campus. By comparing AQI values calculated from the low-cost sensors to those calculated from regulatory monitoring stations, it was determined that the low-cost sensors could effectively measure the air quality that PGSS students were exposed to throughout the program.

## I. Introduction

### A. Common Air Quality Terms and Compounds

Air quality is the level of pollutant concentration in the air and how clean the air is for humans or the environment. Five major air pollutants include particulate matter (PM), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>), and ozone (O<sub>3</sub>).<sup>1</sup> The most important of these pollutants are PM<sub>2.5</sub>, O<sub>3</sub>, and CO<sub>2</sub>. But what are these pollutants?

Particulate matter is a collection of liquid and solid particles in the air that can have varying chemical compositions.<sup>2</sup> Particulate matter can be visible to the human eye or only visible through an electron microscope.<sup>2</sup> The physical size range of particulate matter can range from 2.5 to 10 micrometers in diameter. PM<sub>2.5</sub> is classified as any particle that is equal to or smaller than 2.5 micrometers in diameter. PM can also come from dust, dirt, or harmful emission sources such as chemical plants. They may also naturally form in the atmosphere. PM can become quite harmful when it is 2.5 micrometers or smaller in diameter. Particles smaller than 2.5 micrometers can become lodged deep inside the lungs and stuck inside the body. A dust particle can range from size 1-400 micrometers in diameter.<sup>2,3</sup>

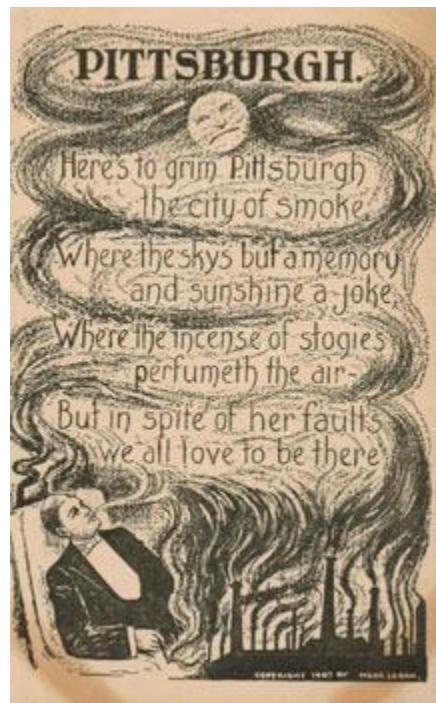
Carbon dioxide is emitted from burning fossil fuels, humans exhaling, and natural processes such as wildfires. Though CO<sub>2</sub> is a minimally toxic air pollutant, it is a major greenhouse gas contributing to climate change.<sup>4,5,6</sup> It is naturally found in the atmosphere in safe quantities. However, high concentrations of CO<sub>2</sub> can cause some limited negative health effects.

Ground-level ozone is an additional example of an airborne pollutant. Atmospheric reactions of pollutants like CO<sub>2</sub> form ground-level ozone. Ground-level ozone is known to inhibit plants' uptake of oxygen.<sup>7</sup> PM, CO<sub>2</sub>, and O<sub>3</sub> are the world's most common and hazardous pollutants.

## B. The Historical Background of Air Quality in Pittsburgh

Air quality events has been prevalent for a very long time around the globe. The earliest record of air quality disasters can be traced to wildfires, volcanic eruptions, and other natural disasters. In addition, air quality events have only increased with human inventions and the mass displacement of waste. Prime examples of how humans have artificially caused air quality events can be observed in the state of Pittsburgh's air quality.

In 1762, a coal seam was found along the Monongahela River.<sup>8</sup> Since the river ran through Pittsburgh, it was natural that Pittsburgh would become a coal mining hotspot. America was hungry for industrialization. Thus, Pittsburgh's air quality and general health conditions would pay the price. Along with coal mining, there was also coal burning and, consequently, coal smoke emissions. When coal is burned, it produces highly toxic PM<sub>2.5</sub> that can contain heavy metals such as mercury, lead, and sulfur dioxide. Furthermore, coal smoke has been known to release carcinogens into the air, earth, and water. These negative effects are precisely what would happen in Pittsburgh, and as coal mining increased, so did the amount of coal smoke. Soon after coal mining became well-established, Pittsburgh would often be covered in heavy smoke and thus was called "The City of Smoke." Many citizens would take notice of the smoke and accept it as a part of life and even as a sign of prosperity. Some citizens even mocked the situation, such as in the poem, *Pittsburgh*, published in 1907 by Meda Logan (See Figure 1 for Meda Logan's poem).<sup>9</sup>



**Figure 1: Pittsburgh by Meda Logan<sup>9</sup>**

Due to the general tolerance of Pittsburgh's heavy smoke, the air quality issue in Pittsburgh persisted. Coal smoke emissions would only increase as, in 1875, Andrew Carnegie turned his interest to steel production and built his first steel mill, J. Edgar Thomson Steel Works, right outside Pittsburgh. Carnegie's steel mill

would be the catalyst for Pittsburgh's steel production and the continuation of coal smoke emissions.<sup>10</sup> For reference to how damaging a steel mill can be, the World Steel Association reported in 2020 that for every ton of steel produced, 1.89 tonnes of CO<sub>2</sub> is emitted into the atmosphere, which is nearly 4,200 pounds of CO<sub>2</sub>.<sup>11</sup> In the 1940s, long after adopting steel mills, Pittsburgh became America's epicenter of steel production. The mills and mines would only worsen the air quality of Pittsburgh and have its name change from "The City of Smoke" to the "City of Steel" or "Hell With the Lid Off." By the 1940s, Pittsburgh's coal smoke pollution became a great interest to the public as it was a massive hindrance to life. Pittsburgh would be burdened with cleaning a constant build-up of coal soot, polluted water from coal and sewage waste directly to the river, a typhoid epidemic, and even soil being unable to grow mostly anything.<sup>12</sup> Pittsburgh citizens even claimed that the smoke would sometimes blot out the sun for hours.<sup>13</sup> The Pittsburgh citizens' claims are promptly supported by photographs between 1940 and 1950 from the Smoke Control Lantern Slide Collection from the Archives Service Center, University of Pittsburgh (See figures 2,3 and 4 for dubious Pittsburgh air quality).<sup>13,14</sup>



**Figure 2: Pittsburgh Smog Blocks The Sun<sup>14</sup>**



**Figure 3: A Common Street In Pittsburgh<sup>14</sup>**



**Figure 4: Pittsburgh Sanitation Cleaning Coal Soot<sup>14</sup>**

For quite some time, Pittsburgh could ignore air pollution because coal mining and steel production was profitable. However, the effects on the city and its citizens were too dire to ignore. Consequently, in 1941, after years of failed attempts at smoke control, Pittsburgh enforced the Smoke Control Ordinance to regulate fuel types and fuel-burning technology. The Smoke Control Ordinance was adequate.<sup>15</sup> From 1945 to 1950, more than half of Pittsburgh's households switched from using coal to natural gas, which ultimately aided in reducing Pittsburgh's coal smoke pollution and improving air quality.<sup>16</sup>

Similarly, the United States Federal Government recognized that many states other than Pennsylvania also suffered from air quality issues and took minor action. However, it was not until The Donora Smog that the Federal Government decided to take things seriously and pass the clean air act.

In the 1940s, the United States Federal Government recognized air pollution as a severe issue, with the first significant air pollution disaster in America, the 1948 Donora Pennsylvania Smog. The disastrous smog lasted from October 27 to October 30, 1948. The air pollution killed 20 people and poisoned nearly half of the town.<sup>17</sup> The specific cause is not entirely known, but it is plausible that a significant contributing factor was the pollutants emitted by American Steel and Wire plant and Donora Zinc Works.<sup>17</sup>

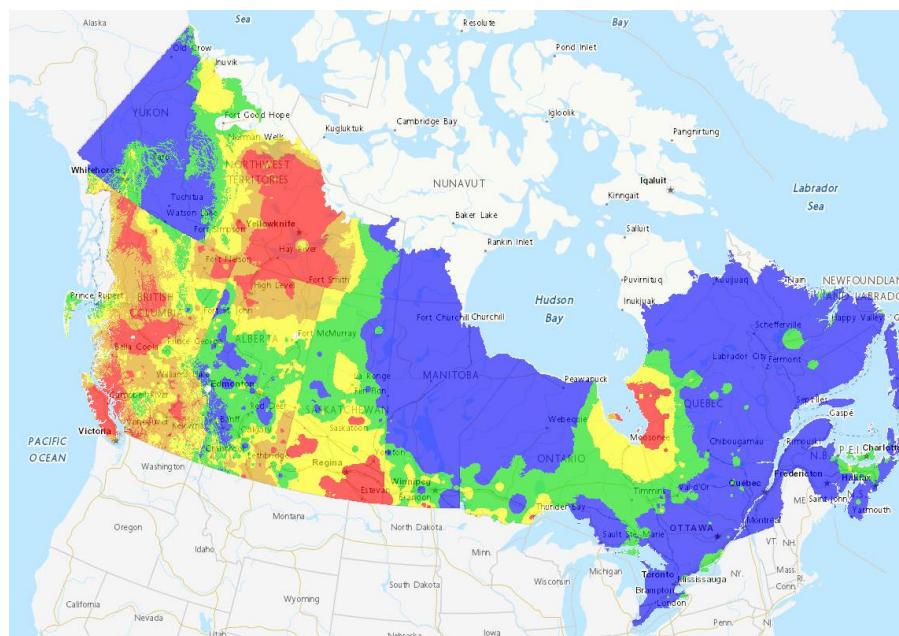
Due to places like Pittsburgh and disasters like the Donora Smog event occurring in the United States, the Federal Government took action to promote clean-air legislation. On December 2, 1970, the Environmental Protection Agency (EPA) was created, and the Clean Air Act of December 31, 1970, was established. Each of these actions worked towards the enforcement of clean-air legislation. The Clean Air Act of 1970 sanctioned federal and state regulations limiting stationary and mobile emissions sources.<sup>18</sup> Furthermore, the EPA enforced the clean air act, so unlike previous attempts, the Clean Air Act would have an active enforcer. With Pittsburgh and the federal government recognizing the degradation of air quality, both parties were able to reduce emissions and return not only Pittsburgh but other cities and states back to healthier air quality and living conditions.

From the 1940s to the 1970s, America has a more analytical mindset regarding the environment. More specifically, maintaining air quality to prevent events like Pittsburgh or Denora from reoccurring. As proof of that, in 1999, the EPA created the Air Quality Index, which has been updated to reflect the times latest health standards.<sup>19,20</sup> Despite all of the changes made by the Federal Government and Pittsburgh, there are still unfortunate side effects that still follow Pittsburgh's residents in the form of Asthma.

The poor air quality in Pittsburgh leads to a higher rate of asthma and other respiratory illnesses. In a recent study of 1200 students in Allegheny County, 22% had asthma.<sup>21</sup> The national average for childhood asthma is 8%.<sup>21</sup> This difference (a 175% increase over the national average) can be attributed to Pittsburgh's higher rates of industrial processes compared to other cities in the United States. Additionally, the American Allergy and Asthma Foundation ranked Pittsburgh the 53rd most challenging city to live in with asthma out of 100 of the largest American cities.<sup>22</sup> The statistics mean that Pittsburgh not only has a shockingly high asthma rate, and those with asthma experience challenging air quality that may worsen their symptoms.

### C. Canadian Wildfires

In April of 2023, a major wildfire started in Canada. By early May, this fire started a series of smaller, severe wildfires in Pittsburgh. These fires were mainly located in the eastern regions, such as Ontario and the Western provinces of Alberta and British Columbia. By early July, the fires had spread to Canada's Northwest Territories. (See Figure 5 for a map of the Canadian wildfires<sup>24</sup>) These fires have continued into early August. The Canadian wildfires of 2023 have burned over 27 million acres and displaced over one hundred thousand residents across Canada.<sup>23</sup>



**Figure 5: Fire Map of Canada on July 15, 2023<sup>24</sup>**

The air quality result of these forest fires is the production of copious amounts of dangerous fine particulate matter. The most common byproduct of these fires is black carbon. Black carbon is one of the atmosphere's most abundant forms of PM<sub>2.5</sub> in the atmosphere. The combustion or burning of biomass primarily creates it since it does not burn cleanly and releases lone carbon molecules forming small chains. These chains can be carcinogenic and cause various other negative health effects. Black carbon is also emitted by burning gasoline and other hydrocarbon-based fuels. However, since these fuel sources are far more uniform than biomass, they tend to burn cleaner. Black carbon is also known to be a primary pollutant driving climate change.

## D. Health Effects of Poor Air Quality

Of the many types of environmental toxins and contaminants that harbor negative health effects, particulate matter is one of the most important and common ones. Particulate matter may enter the respiratory system and cause minor to life-threatening long-term consequences. Of the common types of particulate matter, PM<sub>2.5</sub> and PM<sub>10</sub>, PM<sub>2.5</sub> are more closely related to illness and health degradation. The smaller particles in PM<sub>2.5</sub> can travel further into the lungs and, eventually, the bloodstream than the larger particulate matter, making it more harmful. Air contaminated with high quantities of PM<sub>2.5</sub> has been linked to widespread illness in industrial cities.<sup>25</sup>

Particulate matter is known to result in damage to the entire body system, including most organs.<sup>25</sup> High exposure to particulate matter over a long period has been linked to acute cardiac problems such as arrhythmia, myocardial infarction (heart attack), myocardial ischemia, and heart failure. Particulate matter exposure increases the likelihood of acute, sudden, decreased blood flow to the heart and strokes. Additionally, it has a solid connection to asthma and bronchitis prevalence, especially in children. European Study of Cohorts for Air Pollution Effects (ESCAPE) research demonstrated a connection between PM<sub>2.5</sub> exposure and lung cancer,<sup>27</sup> strokes,<sup>28</sup> and heart failure.<sup>29</sup> Moderate exposure may cause respiratory inflammation. Prolonged exposure damages all organs and may lead to premature death.<sup>25</sup> The EEA estimated in 2019 that over 340,000 premature deaths occurred due to particulate matter.<sup>25,26</sup>

Recent studies on the effects of particulate matter on viral disease spread during the COVID-19 pandemic suggest a strong correlation between PM<sub>2.5</sub> exposure and disease severity. PM<sub>2.5</sub> weakens the immune system and the respiratory systems, allowing diseases to spread more easily and harmfully in polluted areas. Persons experiencing respiratory inflammation would be more prone to severe infection or mortality in the case of a viral infection such as COVID-19. Poor air quality induces an immuno-compromised state, making such illnesses more severe and life-threatening. This correlation was observed in the COVID-19 pandemic as members of more industrial, polluted areas contracted more severe cases and exhibited a higher mortality rate.<sup>25</sup>

Two other gases that are important to monitor related to air quality and health are ozone (O<sub>3</sub>) and CO<sub>2</sub>. Both O<sub>3</sub> and CO<sub>2</sub> are naturally occurring in the atmosphere and do not cause any major negative health effects in their natural quantities. However, increased concentrations of these gases are a cause for concern.

The inhalation of O<sub>3</sub> results in respiratory irritation and inflammation, congestion, chest pain, and coughing. Other health effects include worsening of asthma and bronchitis. According to the National Institute of Occupational Safety and Health and OSHA regulations, the maximum non-hazardous concentration of O<sub>3</sub> for an extended period is 0.1 parts per million or 100 parts per billion. Anything greater than 100 parts per billion is considered hazardous and known to affect respiratory health negatively. Additionally, increased O<sub>3</sub> inhalation has been linked to premature death.<sup>30</sup>

CO<sub>2</sub> is a common gas that makes up a sizable portion of the atmosphere. Before the industrial revolution, the earth's atmosphere comprised between 280 and 350 parts per million CO<sub>2</sub>. Today the atmosphere is composed of roughly 420 parts per million CO<sub>2</sub>, which is a sizable increase, but still not unhealthy. The range of 400 to 1000 parts per million CO<sub>2</sub> is deemed healthy and non-irritating, with no adverse health effects attributed to it. Outdoor CO<sub>2</sub> levels rarely exceed 1000 parts per million other than in industrialized areas with heavy emissions. The primary concern regarding CO<sub>2</sub> is indoor concentration since respiration produces carbon dioxide. If many people are occupying a poorly-ventilated area, such as a small room, the CO<sub>2</sub> concentration can easily reach unsafe levels.<sup>31</sup>

OSHA sets the limit for 8-hour CO<sub>2</sub> exposure at 5000 parts per million.<sup>32</sup> At around 5000 parts per million of CO<sub>2</sub>, many people tend to experience drowsiness and other mild symptoms. Additionally, long-term exposure is known to cause mild cognitive decline. To minimize CO<sub>2</sub> exposure, it is recommended that homeowners update their HVAC systems if there are more than 2000 parts per million CO<sub>2</sub> in the air in their houses.<sup>33</sup> One important metric for measuring the safety of a building is its air exchange rate. This describes how often a room's air is circulated over time. The higher the air exchange rate, the faster CO<sub>2</sub> can be removed and the safer the building is for extended periods. Poor air exchange rates are associated with the negative symptoms of excessive CO<sub>2</sub> inhalation, such as drowsiness or dizziness.

## E. Air Quality Sensors

Air quality sensors are used to measure the concentrations of pollutants within the air. These devices collect data, then transmit it to a database and or store it on an SD card. This information is beneficial when it comes to studying the contents of the air and determining whether the air is safe to breathe in. Air quality sensors are extremely useful in analyzing air pollution history, current air pollution patterns, and possible future air pollution increases or decreases.

### 1. High-Cost Regulatory Monitoring Station Sensors

The data used by the United States government to calculate daily air quality indexes (AQIs) originates from many different air quality sensors stationed around the country. The Environmental Protection Agency (EPA) mainly uses the beta attenuation mass (BAM) system to monitor PM<sub>2.5</sub> levels. Through this system, beta particles are passed through particulate matter that is collected on a filter.<sup>34</sup> The amount of beta particles that are not absorbed by the particulate matter is quantified and compared to the original number of beta particles released.<sup>34</sup> The system can then determine how much particulate matter has been collected from the air by calculating how many beta particles were absorbed.<sup>34</sup> This process uses carbon-14, a radioactive isotope of carbon, that has an energy level sufficient to release the number of beta particles needed for measurement while being relatively safe to work with.<sup>34</sup> Regarding its high-cost sensors, the EPA follows the federal reference method. This method is a testing strategy used to ensure all regulatory monitoring station sensors measure pollutant concentrations similarly. When sensors are running this way, they are a part of the reference method.<sup>35</sup> The sensors used by the EPA to take measurements can cost anywhere from \$20,000 to \$100,000.<sup>36</sup>

### 2. Low-Cost Particle Sensors

Low-cost air quality sensors are typically much smaller than the sensors used by the government and use less advanced technology. This paper focuses on using real-time, affordable, multi-pollutant (RAMP) sensors. These sensors were developed by Carnegie Mellon's Center for Atmospheric Particle Studies, and SenSevere, a company dedicated to producing sensors that will help keep people safe when working in severely dangerous environments.<sup>37</sup> As opposed to the BAM system used by regulatory monitoring station sensors, RAMP sensors use a light scattering technique to measure PM<sub>2.5</sub> concentrations.<sup>38</sup> This technology comprises an infrared emitting diode, a phototransistor, and a focusing lens. When a sensor collects PM<sub>2.5</sub> particles through a cyclone, the particles enter the beam of light within the sensor and scatter it. The phototransistor measures the intensity of the scattered light. The measured intensity is then correlated to the concentration of PM<sub>2.5</sub> in the air sample collected. This technology is used as opposed to BAM systems because it is much cheaper to manufacture and run and requires much less advanced technology and power supply. The response time of the light scattering technology is also very short, which means sensors using this technology can send feedback to a database very often<sup>38</sup>. In this experiment, data were sent once every minute. Rapid response time is desirable when achieving highly precise measurements and large amounts of data.

### 3. Low-Cost Gas-Phase Sensors

Passive sensors are used to monitor the concentrations of gas-phase chemical species. An example of such a passive sensor is an electrochemical sensor. The RAMP sensors used in this experiment can measure CO, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO<sub>2</sub> concentrations. A nondispersive infrared (NDIR) sensor was used to measure CO and CO<sub>2</sub> concentrations, as well as temperature and relative humidity, which are also crucial pieces of data when forming conclusions about PM<sub>2.5</sub> concentration data.<sup>39</sup> The remaining compounds, SO<sub>2</sub>, NO<sub>2</sub>, and O<sub>3</sub>, were measured by electrochemical sensors. The process by which electrochemical sensors work relies on a working electrode and a counter electrode. When the air being measured reaches the working electrode, an oxidation or reduction reaction occurs. This reaction causes a shift in electrons which generates an electrical current. The magnitude of the electrical current correlates to the concentration of the gas being detected. An additional electrode is used for measuring temperature and relative humidity.<sup>39</sup> Cross sensitivity can interfere with the measurements taken by electrochemical sensors. Cross sensitivity occurs when a sensor detects the presence of the gas it is looking to measure but is not actually present. It is instead measuring a different gas entirely. Cross sensitivity can throw off concentration readings when electrochemical sensors report skewed data after they factor the incorrect compound into their measurements.<sup>40</sup> Because of the possibility of cross-sensitivity or other environmental interferences, careful corrections need to be applied to the data collected from the low-cost gas-phase sensors to produce data that is of the same quality as the data collected by regulatory monitoring stations.

### 4. Motivation

High-cost regulatory monitoring station sensors are an effective tool for measuring a given area's general air quality conditions of a given area. However, they cannot detect small-scale events that may cause air quality to differ within a smaller radius. Because of this, low-cost sensors are necessary for those who want to know the precise air quality measurements of the neighborhood or town they live in. For some people, precise air quality is a necessity, as they may suffer from lung conditions, heart conditions, or severe allergies and may be impacted by any air quality that is less than exceptional. Because towns and cities with populations less than 350,000 people are not required to report an AQI (air quality index), this valuable data may not be available to those who need it.<sup>41</sup> It would not be a viable option to add enough regulatory monitoring stations to produce such accurate measurements, as such sensors are costly and quite difficult to implement due to their size and complex technical components. Low-cost sensors, in comparison, are inexpensive, small, and easy to implement. Those interested in learning more about the air quality near their home could set up a low-cost RAMP sensor in their yard or on their balcony, roof, or windowsill. By placing these sensors around their living space, people could see an accurate depiction of the air they are exposed to often rather than a depiction of the air a mile away. These low-cost sensors could also be used inside a home as well due to their relatively small size. This is an important feature, as most of the population spends most of their time indoors, and knowing exactly what compounds they are exposed to is important for people with sensitivities. This fact remains true at the Pennsylvania Governor's School for the Sciences. Students spend most of their time indoors attending lectures, completing homework, eating, and sleeping. This paper analyzes the data collected by low-cost sensors to determine if they can be used to accurately measure and report air quality exposure faced by students during the PGSS program.

## II. Methods

### A. PM<sub>2.5</sub> Corrections

The PM<sub>2.5</sub> measurements taken by the sensors needed to undergo some corrections before they were ready to be used for data analysis. The formulas used to complete such corrections were derived from a paper published by Malings et al. (2019),<sup>42</sup> which focuses on what corrections need to be made to low cost sensor data in order to make it comparable to regulatory monitoring station sensors, also known as BAM-equivalent.

#### 1. Water Activity Correction

Equation 1 was used to account for water activity within the air. Water activity is a measure of the energy status of the water in a given system. The water activity equation that was used to correct the data factors in the changing temperature and relative humidity conditions seen throughout the experimental period.<sup>42</sup>

$$a_w(T, RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w R T D_p}\right)^{-1} \quad (1)$$

In Equation 1,  $\sigma_w$  represents the surface tension of water,  $M_w$  represents the molecular weight of water, and  $\rho_w$  represents the density of water;  $T$  is the absolute temperature,  $R$  is the ideal gas constant,  $RH$  is the ambient relative humidity; and  $D_p$  is the diameter of the particle being corrected.<sup>42</sup>

#### 2. Hygroscopic Growth Correction

Only correcting the data for water activity was not sufficient. The relative humidity (RH) alone has a very large impact on the PM<sub>2.5</sub> readings taken by the sensors. This is because of hygroscopic growth,<sup>44</sup> which occurs when fine airborne particles are exposed to varying humidity levels. At higher humidity levels, particles will absorb water, and at lower humidity levels, they will release water. These changes cause the size of the particles to change, which throws off the sensors, as they are specifically designed to measure particulate matter with a diameter of 2.5 micrometers or less. Equation 2 was used to correct for this interference. To account for the changes in humidity, a ratio was created between the PM<sub>2.5</sub> measurement at a varia. However humidity (RH) and temperature (T) to a PM<sub>2.5</sub> measurement at 22°C and 35% humidity. These conditions are the standards set by the EPA for data collection.<sup>42</sup>

$$fRH(T, RH) = 1 + \kappa_{bulk} \left( \frac{a_w(T, RH)}{1 - a_w(T, RH)} \right) \quad (2)$$

Within Equation 2,  $a_w$  represents the water activity calculated in equation 1, and  $\kappa_{bulk}$  represents the hygroscopicity of bulk aerosol.  $\kappa_{bulk}$  changes depending on the season due to changes in air particle composition.<sup>44</sup> By evaluating these changes specific to Pittsburgh, the winter  $\kappa_{bulk}$  value determined was 0.34, and the summer  $\kappa_{bulk}$  value determined was 0.26. Through this experiment, the value of 0.26 was used, as the data was collected throughout the month of July.<sup>42</sup> These values were derived by a paper written by Malings et al. (2019).

#### 3. Aerosol Calibration Correction

Even after factoring in changes caused by relative humidity, there was still a correction that needed to be done to account for the fact that the aerosols the sensors are calibrated with within the factory are different from those in the atmosphere in Pittsburgh. Equation 3 is used to achieve such results.

$$[\text{corrected } PM_{2.5}] = \theta_1 \left( \frac{[PM_{2.5} \text{ as reported}]}{fRH(T, RH)} \right) + \theta_0 \quad (3)$$

In Equation 3, fRH represents the hygroscopic growth ratio calculated using Equation 2,  $\theta_0$  and  $\theta_1$  are coefficients determined through linear regression techniques in this experiment, the value of  $\theta_0$  used was 5.28, and the value of  $\theta_1$  used was 1.50. Both the linear regression and techniques and variable values were derived from the Malings et al. (2019) paper that describes the corrections required for data measured by low-cost sensors.<sup>42</sup>

Equations 1, 2, and 3 were ultimately applied to the collected PM<sub>2.5</sub> data spreadsheet. This meant that the data collected by the low-cost sensors were comparable to the data collected by the stations that operate under EPA standards. In other words, the low-cost data was corrected to become BAM-equivalent.

## B. Air Quality Index Calculations

### 1. What is the Air Quality Index?

The air quality index (AQI) is a way to summarize the concentrations of pollutants in the atmosphere and what they mean for the people in each area. An AQI is calculated by converting a pollutant's concentration into a uniform index. Originally developed by the EPA with the purpose of simplifying air pollution readings for the public, the AQI is used globally to report air quality conditions to the general public who may not understand raw pollutant concentration data.<sup>43</sup> The scale of AQI is reported using a scale that ranges from 0 to 500. Table 1 demonstrates the way the range is broken down to describe certain health impacts for any given AQI.

**Table 1: Air Quality Index Ranges**

AQI Range	Descriptor
0 to 50	Good
51 to 100	Moderate
101 to 150	Unhealthy for Sensitive Groups
151 to 200	Unhealthy
201 to 300	Very Unhealthy
301 to 500	Hazardous

As demonstrated by Table 1, an AQI of 0 to 50 is good. At this level there are no risks associated with the air quality. An AQI within the range of 51 to 100 is considered moderate, which means that some pollutant concentrations may be of slight concern to those who are unusually sensitive to air pollution. When AQI is within the range of 101 to 150, it is considered unhealthy for sensitive groups, meaning that people with heart or lung disease, elderly people, and young children may experience some health effects. When the AQI falls in the range of 151 to 200, it is considered unhealthy to all people, meaning that the general public may face health effects, and sensitive groups may face serious effects. An AQI between 201 and 300 is

considered very unhealthy, meaning that all people may experience serious health effects. In the final range, 301 to 500, conditions are considered an emergency, and the entire population is very likely to be impacted by the air pollution. These ranges are helpful for the public because they simplify the data collected by air quality measuring stations and tell the general public what they truly need to know: whether their health is at risk or not.<sup>44</sup>

## 2. How is AQI Calculated?

Typically, when the AQI is reported, it will be based on the pollutant with the highest AQI so that the public is fully aware of any possible danger.<sup>45</sup> It is possible, however, to narrow an AQI search in order to evaluate the AQI based on any pollutant. The AQI pollutant of interest in this paper is PM<sub>2.5</sub>. The PM<sub>2.5</sub> concentration data collected from the low-cost sensors was compared to the data collected by regulatory monitoring station sensors, but only after it was converted to the AQI, as this will allow for discussion as to if the low-cost sensors can indeed be used to monitor air quality and possible health effects during PGSS. Equation 4 represents the formula used to convert raw concentration data to AQI values.

$$I_p = ((I_{Hi} - I_{Lo}) / (BP_{Hi} - BP_{Lo})) (C_p - BP_{Lo}) + I_{Lo} \quad (4)$$

Within Equation 4, C<sub>p</sub> represents the measured PM<sub>2.5</sub> concentration truncated to one decimal place, BP<sub>Hi</sub> represents the breakpoint that is greater than C<sub>p</sub>, and BP<sub>Lo</sub> represents the breakpoint that is lower than C<sub>p</sub>. I<sub>Hi</sub> is the AQI value that corresponds to BP<sub>Hi</sub> and I<sub>Lo</sub> is the AQI value that corresponds to BP<sub>Lo</sub>. The breakpoints can be correlated to the AQI value using Table 2.<sup>41</sup>

**Table 2: PM<sub>2.5</sub> Breakpoints and Correlating AQI**

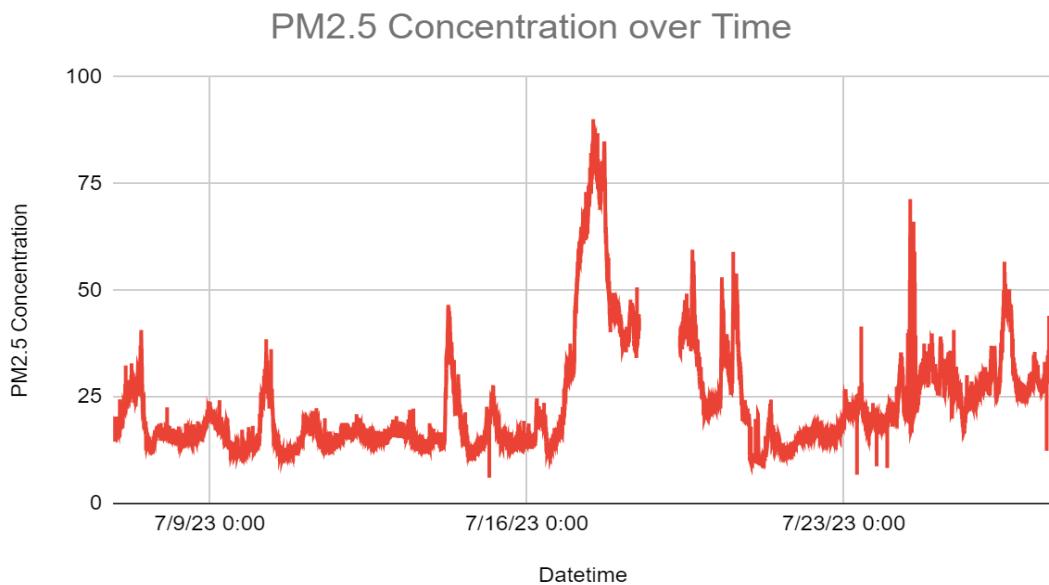
PM <sub>2.5</sub> Breakpoints (µg/m <sup>3</sup> )	Correlating PM <sub>2.5</sub> AQI
0.0 - 12.0	0 - 50
12.1 - 35.4	51 - 100
35.5 - 55.4	101 - 150
55.5 - 150.4	151 - 200
150.5 - 250.4	201 - 300
250.5 - 350.4	301 - 400
350.5 - 500.4	401 - 500

### III. Results

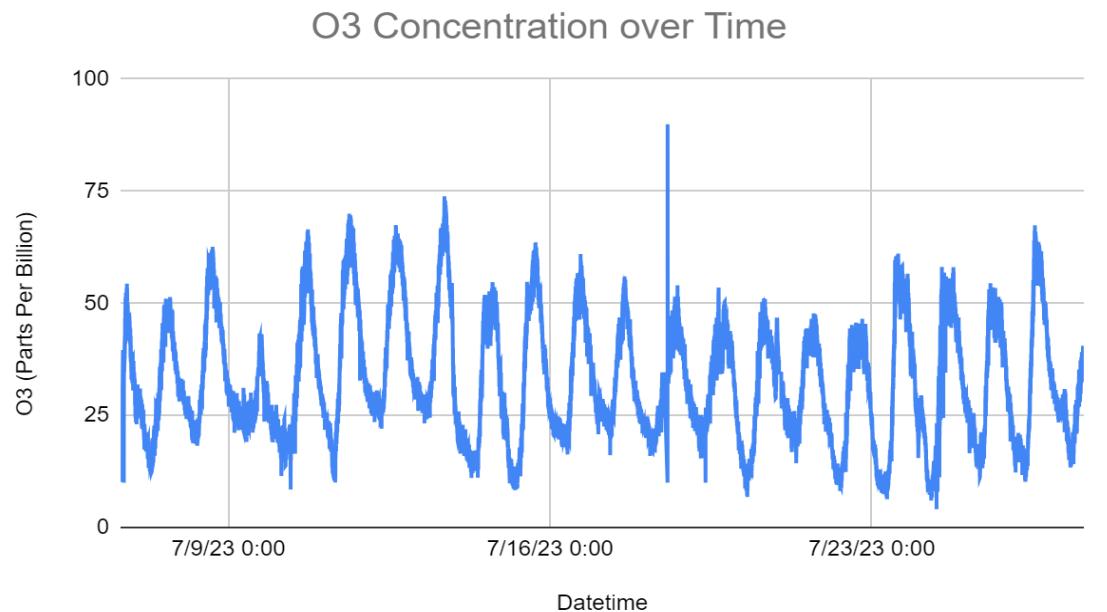
#### A. Overview of Outdoor and Indoor Air Quality

##### 1. Outdoor Air Quality Overview and Analysis

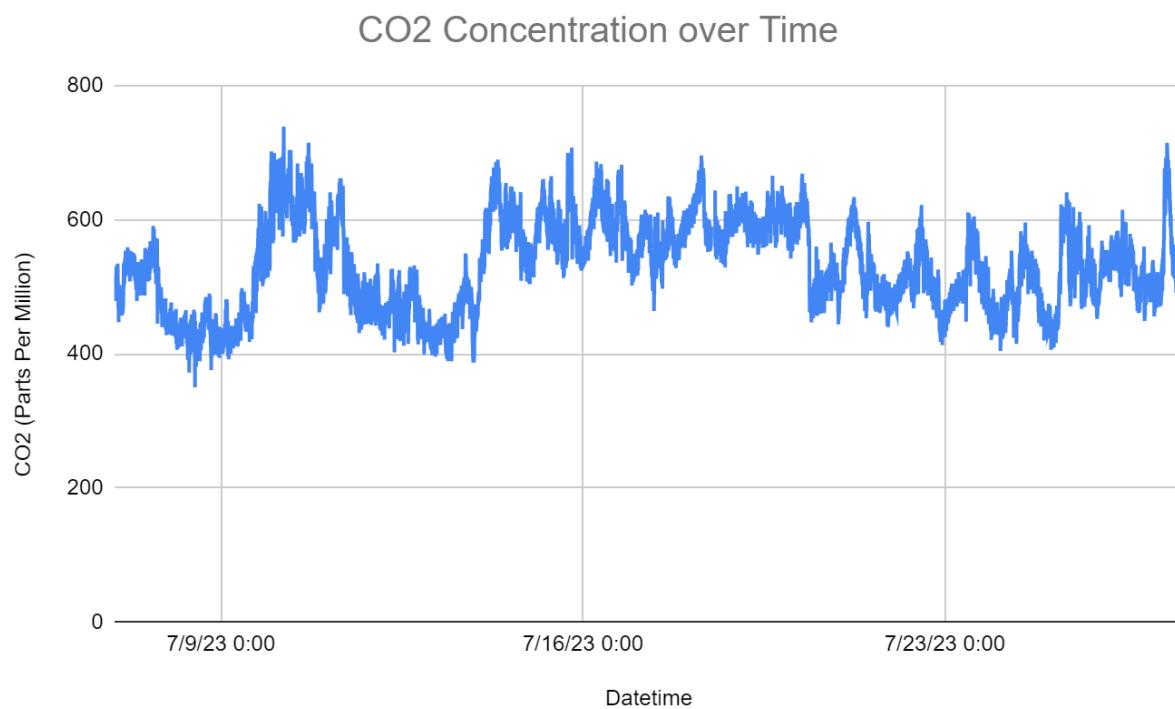
The average outdoor concentration of PM<sub>2.5</sub> over the course of the experiment was 23.14 µg/m<sup>3</sup>. The maximum suggested 24-hour average is 12.5 µg/m<sup>3</sup>. This means that the average concentration recorded during the experimental period was nearly twice the suggested healthy average. (See figure 6 for PM<sub>2.5</sub> concentration outside over time) The average outdoor concentration of O<sub>3</sub> was 32.05 parts per billion (ppb) over the course of the experiment. This is significantly under the 100 ppb limit set by OSHA for 8-hour exposure and the suggested healthy 24-hour average of 70 ppb. The O<sub>3</sub> levels fluctuate on a daily cycle. (For O<sub>3</sub> concentration over time outside, see figure 7) The outdoor CO<sub>2</sub> average was 525.15 parts per million (ppm). (See figure 8 for outdoor CO<sub>2</sub> concentration over time) This concentration is well within the healthy recommended concentration of CO<sub>2</sub> in the air. However, it is significantly higher than the global average of roughly 425 ppm recorded one month earlier. This suggests that Pittsburgh has many major sources of CO<sub>2</sub> and greenhouse gas emissions.



**Figure 6: Outdoor PM<sub>2.5</sub> Concentration over the 20 Sampling Days**



**Figure 7: O<sub>3</sub> Concentration over the 20 Sampling Days**

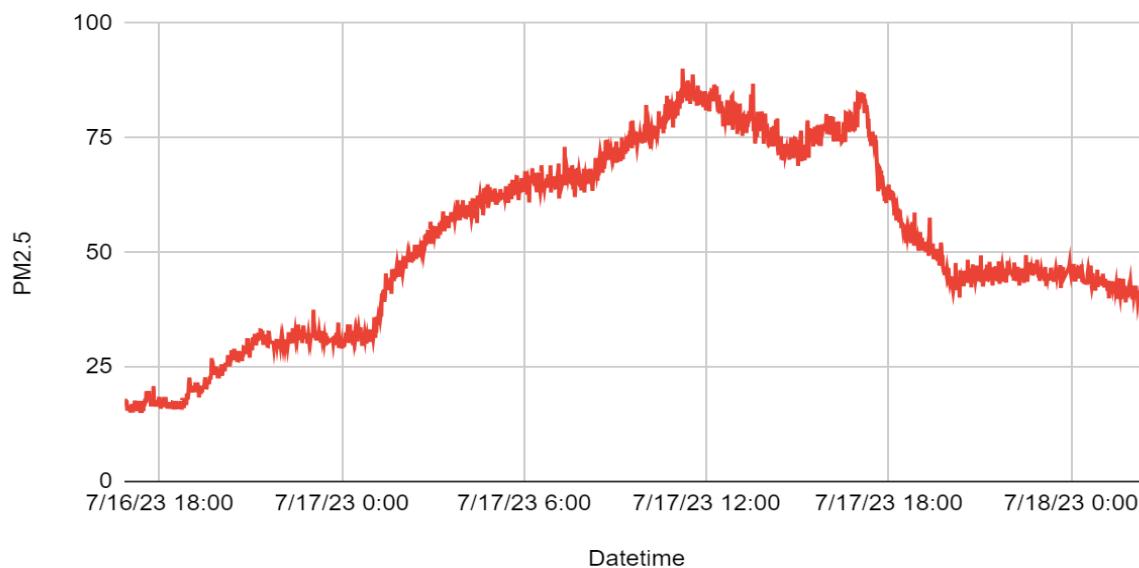


**Figure 8: CO<sub>2</sub> Concentration over the 20 Sampling Days**

On July 17th, there was a spike in PM<sub>2.5</sub> due to wind patterns dispersing smoke from the Canadian wildfire. (See figure 9 for concentrations of PM<sub>2.5</sub> during the air quality event) PM<sub>2.5</sub> concentrations reached a high of 88.82 µg/m<sup>3</sup>, significantly higher than the earlier values recorded, most of which fell below 20. This was far above the suggested maximum of 12.5 µg/m<sup>3</sup>. A concentration of PM as high as 88.82 µg/m<sup>3</sup> poses an

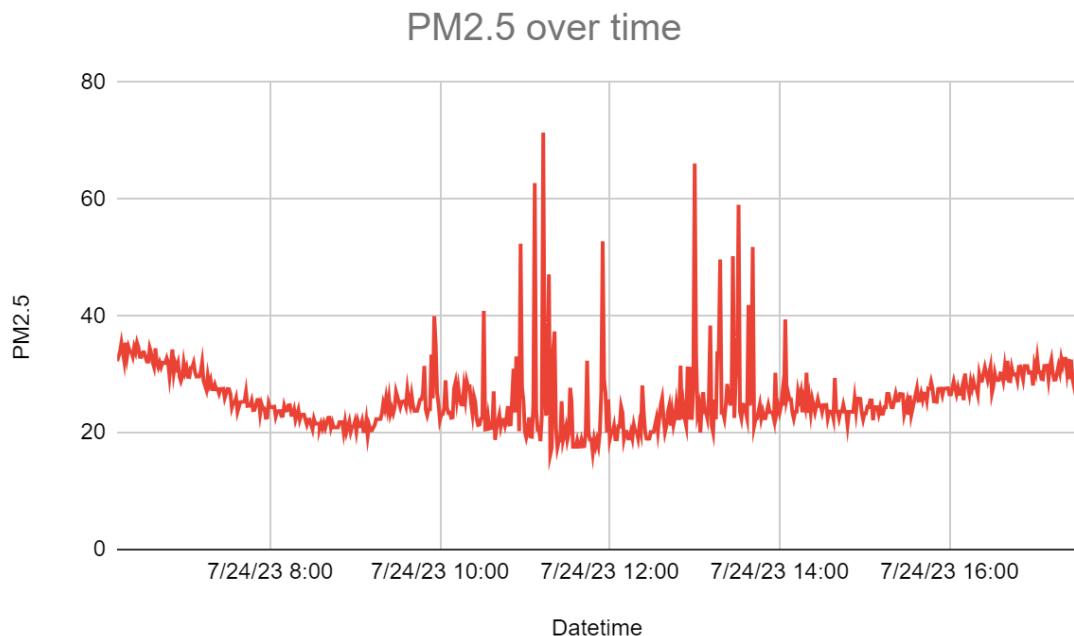
immediate health risk to anyone with asthma, a respiratory condition, or any sort of illness. Extended exposure would likely cause coughing, shortness of breath, sore throat, and respiratory inflammation in otherwise healthy individuals. By the next day, the air quality had significantly improved, but the concentration of PM<sub>2.5</sub> still posed a risk to individuals with respiratory conditions.

### PM2.5 Concentration over Time



**Figure 9: The PM<sub>2.5</sub> Peak on July 17th during Air Quality Event over Time**

Another interesting PM<sub>2.5</sub>-increasing event is the weekly landscaping occurring on Mondays. Every Monday, between 9:00 AM to 3:00 PM, a team of people would cut the grass at Carnegie Mellon University, using large riding mowers and gas-powered weed trimmers. Gas-powered engines and motors are a major source of PM<sub>2.5</sub> and CO<sub>2</sub> emissions. The weekly landscaping created several short-lived spikes of PM<sub>2.5</sub>. Unlike the PM<sub>2.5</sub> pollution from the forest fire, the PM<sub>2.5</sub> produced by landscaping equipment dispersed quickly, reaching equilibrium through diffusion with the nearby, cleaner air. (See figure 10 for an example of landscaping-based PM<sub>2.5</sub> concentration spikes) Since the PM<sub>2.5</sub> spikes produced in this process are so short-lived, it may cause irritation in individuals with respiratory conditions or allergies, but it does not pose any short or long term threat to individuals' health.

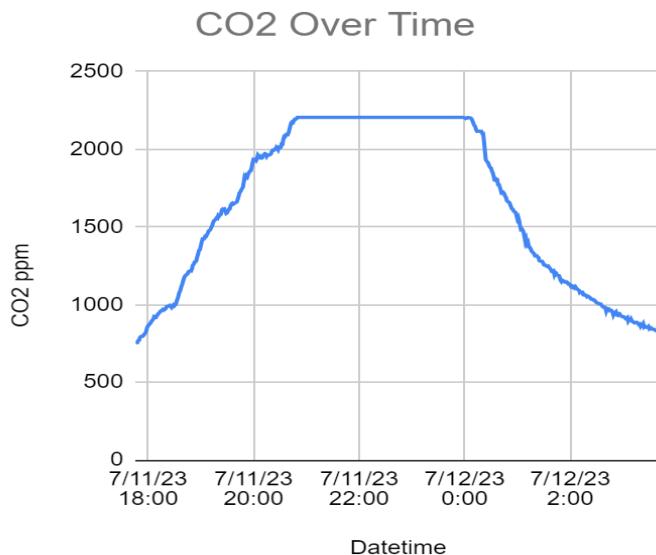


**Figure 10: PM<sub>2.5</sub> Spikes from Landscaping on Mondays over 12 Hour Period**

## 2. Indoor Air Quality Overview and Analysis

The average concentration of PM<sub>2.5</sub> indoors during the experiment was 16.55 µg/m<sup>3</sup>. This is slightly higher than the recommended concentration of 12.5 µg/m<sup>3</sup>. Since it is so close to the healthy, recommended exposure, the indoor concentrations of PM<sub>2.5</sub> do not pose any serious threat to health. It may be slightly irritating to individuals with respiratory conditions and illnesses, however. The average O<sub>3</sub> concentration indoors was 10.045 parts per billion (ppb), which is significantly lower than the upper limit of 100 ppb and the recommended value of 70 ppb. The average value of CO<sub>2</sub> was 987.63 parts per million (ppm). This is significantly higher than the global outdoor average but is not in a range where it would pose a major health risk.

The CO<sub>2</sub> concentrations were not consistent throughout the experiment. Since CO<sub>2</sub> is produced by breath and the sensors were used in areas that hold several people for extended periods of time, there were several CO<sub>2</sub> spikes based on when people stayed in the rooms tested. In the Donner House common area, the CO<sub>2</sub> spiked nightly as students studied in groups. These CO<sub>2</sub> spikes measured as high as 1990 ppm, the maximum value that the sensors can record. With the CO<sub>2</sub> sensor correction conversion value applied, the spikes measured as high as 2205 ppm. The concentrations were likely higher, as many peaks were capped at the maximum value while they would have extended further. (See figure 11 for an example of CO<sub>2</sub> concentrations exceeding maximum sensor values)



**Figure 11: CO<sub>2</sub> Exceeding the Maximum Sensor Value**

The CO<sub>2</sub> concentration peaks are staggeringly high and could even cause mild health effects. CO<sub>2</sub> functions as a mild asphyxiant. In concentrations over 2000 ppm, CO<sub>2</sub> may have a tiring effect on individuals, making them feel drowsy or dizzy. In the common area in Donner House, the CO<sub>2</sub> reached higher than 2205 ppm nearly every night. Students in that area may have experienced these effects.

The second place CO<sub>2</sub> was tested was in a dorm room with an air conditioner, a Dyson Pure Cool air purifier, and two students. At night, the CO<sub>2</sub> levels would increase until they reached over 1600 ppm by the morning before both students left. These CO<sub>2</sub> levels are relatively high compared to the outdoor average, but they are still well within the safe and healthy range of CO<sub>2</sub> concentrations.

The third location for CO<sub>2</sub> testing was another dorm room without an air conditioner or air purifier. One of the windows was open and there were two students occupying the room. The CO<sub>2</sub> levels in this room were lower than in the air-conditioned room because one of the windows was open. In this room, the nightly CO<sub>2</sub> peaks measured approximately 1100 ppm. This is also within the safe range.

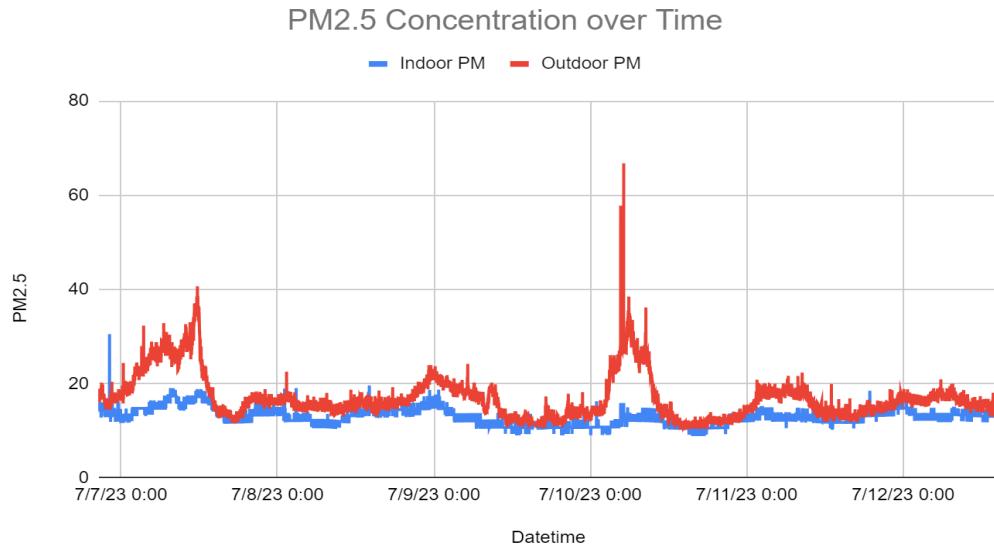
The fourth location where CO<sub>2</sub> was tested was Doherty Hall classroom A302, a lecture hall where most Pennsylvania Governor's School classes take place. Most of the CO<sub>2</sub> spikes had maximum values just above 1000 ppm, though one of its spikes had a maximum value of 1436. This is still within the safe range, and would have little effect on individuals. If students were already drowsy, it may have increased their sleepiness, but it would not have a drastic effect.

## B. Comparison of Indoor and Outdoor Air Quality

### 1. Comparison of Indoor and Outdoor Air Quality in Donner House Common Area

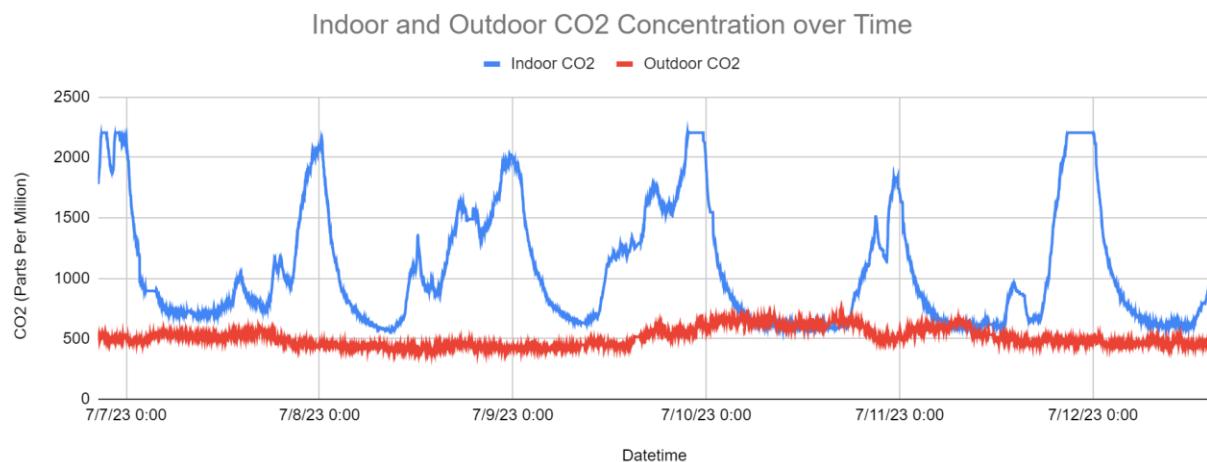
The first indoor location tested for air quality was the Donner House dorm common area. Overall, the common area's air quality was influenced by the fluctuations in outdoor air quality. CO<sub>2</sub> was an outlier, significantly higher inside than outside at strange intervals because of the people inside breathing and creating CO<sub>2</sub>.

The PM<sub>2.5</sub> exposure in the Donner House common area largely followed the trends of PM<sub>2.5</sub> outside. When the particulate matter outside spiked, the indoor particulate matter saw a similar, but much smaller, spike. However, for general trends such as gradual increases in PM<sub>2.5</sub>, the indoor particulate matter followed it much more closely.



**Figure 12: PM<sub>2.5</sub> Concentration over Time Inside Common Area and Outside**

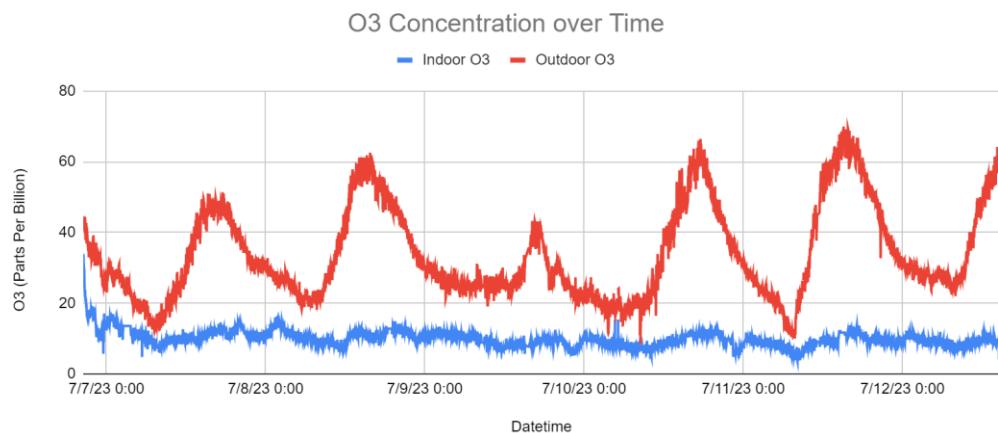
The CO<sub>2</sub> exposure in the Donner House common area was significantly higher than the CO<sub>2</sub> exposure outside. The concentration at night would jump to 2000 parts per million and higher when students were doing their homework and projects at night. The CO<sub>2</sub> concentration in the common area did not follow the same pattern as the CO<sub>2</sub> concentration outside. (See figure 13 for comparison of indoor and outdoor CO<sub>2</sub> concentration) This is likely because the changes in outdoor CO<sub>2</sub> levels are negligible compared to the changes in concentration caused by occupancy and respiration.



**Figure 13: CO<sub>2</sub> Concentration over Time Inside Common Area and Outside**

The O<sub>3</sub> exposure in the Donner House common area was significantly lower than the O<sub>3</sub> exposure outside. However, it did increase and decrease according to the same daily pattern as outdoor O<sub>3</sub>. (See figure 14

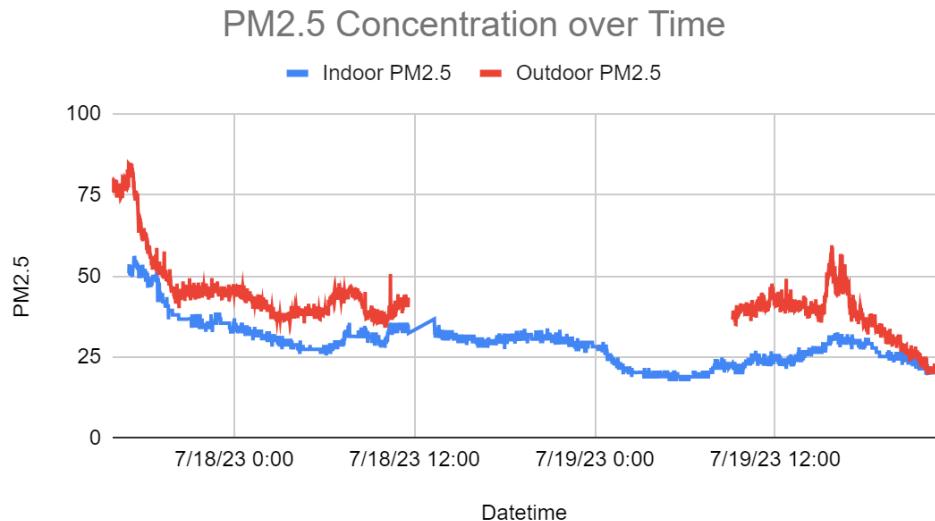
for comparison of outdoor and indoor O<sub>3</sub> concentration) The average value of O<sub>3</sub> over the period of time when the Donner House common area was tested was 32.55 parts per billion. The average inside the Donner House common area was 13.25 ppb. This is a significant decrease, though both averages are well within the safe exposure range.



**Figure 14: O<sub>3</sub> Concentration over Time Inside Common Area and Outside**

The common area was also tested between July 17 and July 19 during an abnormal air quality event. Smoke from the Canadian wildfires had dispersed throughout Pittsburgh due to an unusual wind pattern. This created hazy, poor air. Outside, the PM<sub>2.5</sub> reached a high of 88.8 µg/m<sup>3</sup>, which is significantly above the PM<sub>2.5</sub> exposure limit. The daily air quality index value was 148, which is deemed unhealthy for sensitive groups. During the event itself, there was an hourly air quality index over 150, which is deemed unhealthy. Neither the O<sub>3</sub> or CO<sub>2</sub> concentrations significantly changed compared to the other tests in the common area.

During the event, the PM<sub>2.5</sub> concentration inside also significantly increased. It reached a high of 55.73 µg/m<sup>3</sup>, which could cause respiratory irritation and breathing problems, especially in individuals with respiratory conditions. The indoor PM<sub>2.5</sub> concentration closely followed the outdoor concentration, though it was consistently lower by roughly 10. (See figure 15 for comparison of outdoor and indoor PM<sub>2.5</sub> concentration during the event) Although the air quality inside during that time was also harmful, it was less dangerous than the air outside. The air quality remained poor until after July 19th. During the period of July 17th to July 19th, the outdoor average PM<sub>2.5</sub> concentration was 42.86 µg/m<sup>3</sup> and the average indoor concentration was 29.11 µg/m<sup>3</sup>. Neither of these values are healthy, but it does show that staying indoors during a hazardous air quality event is safer.

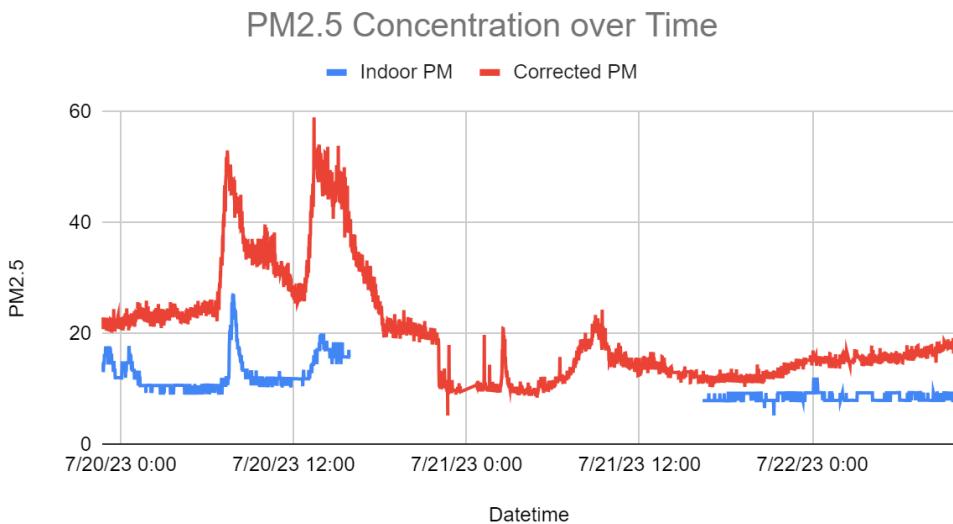


**Figure 15: PM<sub>2.5</sub> Concentration Inside and Outside during Air Quality Event**

## 2. Comparison of Indoor and Outdoor Air Quality in Donner House Dorm Room with Air Purifier

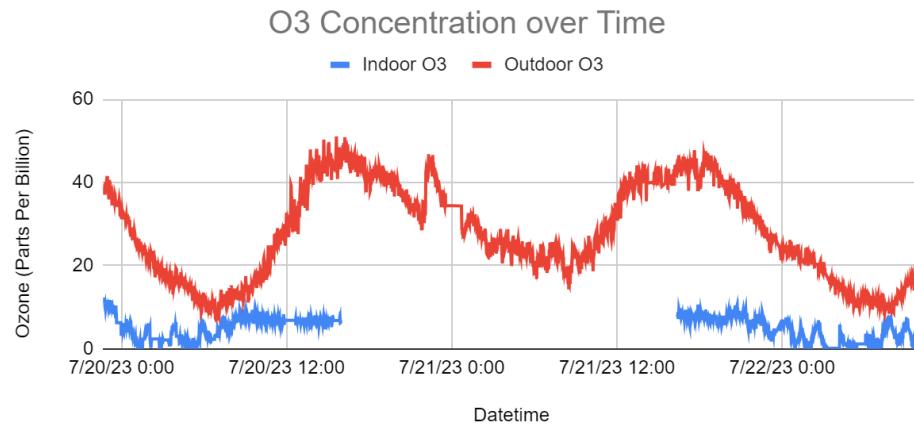
Between 10:40 PM on July 19th and July 22nd at 10:00 AM, the indoor sensors measured the air quality in a dorm room in Donner House with air conditioning and a Dyson Pure Cool air purifier. Two students inhabited this room. Both the PM<sub>2.5</sub> concentration and the O<sub>3</sub> concentration saw a significant decrease indoors. However, like the CO<sub>2</sub> concentration in the common area, the CO<sub>2</sub> concentration in this dorm room saw an increase compared to the CO<sub>2</sub> outdoors.

The PM<sub>2.5</sub> concentration in this dorm room was significantly lower than the concentration outdoors. Inside the room, the concentration was an average of 10.80 µg/m<sup>3</sup>, much lower than the outdoor average of 20.35 µg/m<sup>3</sup>. (See figure 16 for comparison of outdoor and indoor PM<sub>2.5</sub> concentrations over time) This difference is most likely due to the air purifier since it is designed to trap microscopic particles. Unlike the other indoor locations tested, this dorm room met the recommended maximum PM<sub>2.5</sub> concentration of 12.5 µg/m<sup>3</sup>. This suggests that air purifiers have a quantifiable positive impact on air quality.



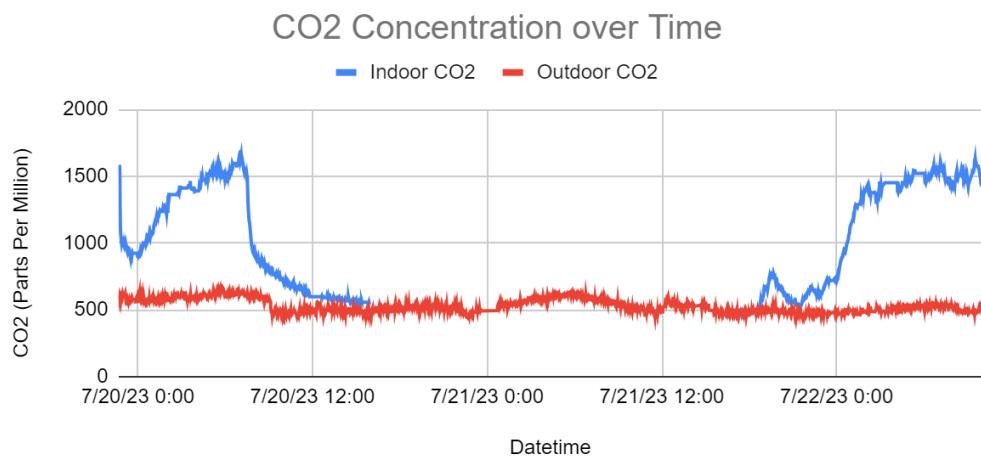
**Figure 16: PM<sub>2.5</sub> Concentration over Time Inside Dorm Room and Outside**

The O<sub>3</sub> concentration saw a highly significant difference outdoors and indoors. The outdoor average concentration over the course of this test was 27.93 parts per billion, which is not a cause for concern. The indoor concentration was 5.04 parts per billion. This is an 82% decrease in O<sub>3</sub> concentration compared to the concentration outside due to the presence of the air purifier. Additionally, the O<sub>3</sub> concentration indoors seemed much less dependent on the daily cycles of the concentration outdoors. (See figure 17 for comparison of indoor and outdoor O<sub>3</sub> concentrations) This is most likely due to the presence of the air purifier in the room. Dyson Pure Cool air purifiers use high efficiency particulate air (HEPA) and carbon filters. Carbon filters are known to decrease the concentration of O<sub>3</sub>.



**Figure 17: O<sub>3</sub> Concentration over Time Inside Dorm Room and Outside**

Much like the common area, the dorm room also exhibited increased CO<sub>2</sub> concentrations. Between the hours of 1:00 AM and 7:00 AM, the CO<sub>2</sub> levels increased dramatically. (See figure 18 for a comparison of indoor and outdoor CO<sub>2</sub> concentration) Unlike in the common area, the CO<sub>2</sub> levels in the dorm room reached a maximum of 1661 and did not exceed the measuring range of the sensor. This range is safer, but could still make sensitive individuals feel drowsy or dizzy.

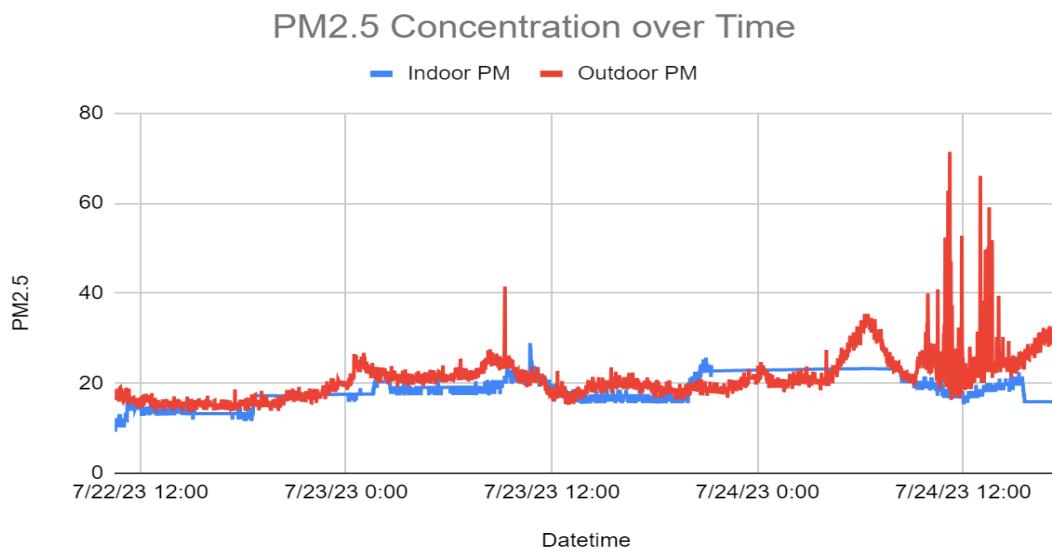


**Figure 18: CO<sub>2</sub> Concentration over Time Inside Dorm Room and Outside**

### 3. Comparison of Indoor and Outdoor Air Quality in Donner House Dorm Room with Open Window

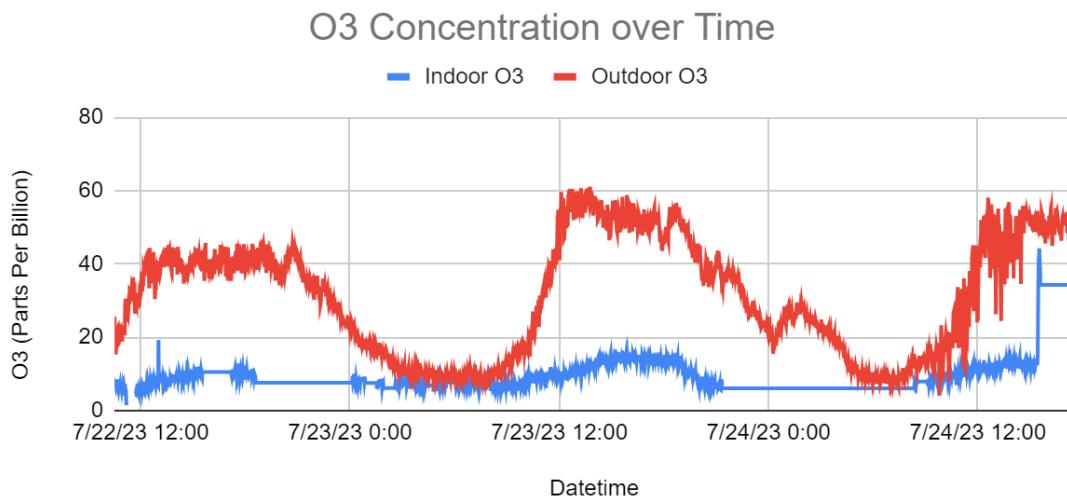
Between 10:30 AM on July 22nd and 5:30 PM on July 24th, the indoor sensor recorded data from another dorm room in Donner House. This dorm room did not have an air conditioner or an air purifier. It had one open window for ventilation and was used by two students. In this dorm room, the PM<sub>2.5</sub>, CO<sub>2</sub>, and O<sub>3</sub> concentrations were much more similar to the concentrations outside.

The PM<sub>2.5</sub> concentration in this dorm was very similar to the concentration outside. The indoor concentration followed the same trends as the outdoor concentration. (See figure 19 for a comparison of the indoor and outdoor PM<sub>2.5</sub> concentration) The only notable difference between the indoor and outdoor concentrations of PM<sub>2.5</sub> occurred between 9:00 AM and 3:00 PM on Monday, July 24. On this day, Carnegie Mellon's campus was undergoing landscaping. The gas-powered engines on lawn mowers and weed trimmers most likely caused the series of short-lived, high peaks in outdoor PM<sub>2.5</sub> concentration. The PM<sub>2.5</sub> concentration in the dorm room did not increase during this period of time, suggesting that the local high concentrations of PM<sub>2.5</sub> dispersed throughout campus without having a significant impact on the air quality on other parts of the campus.



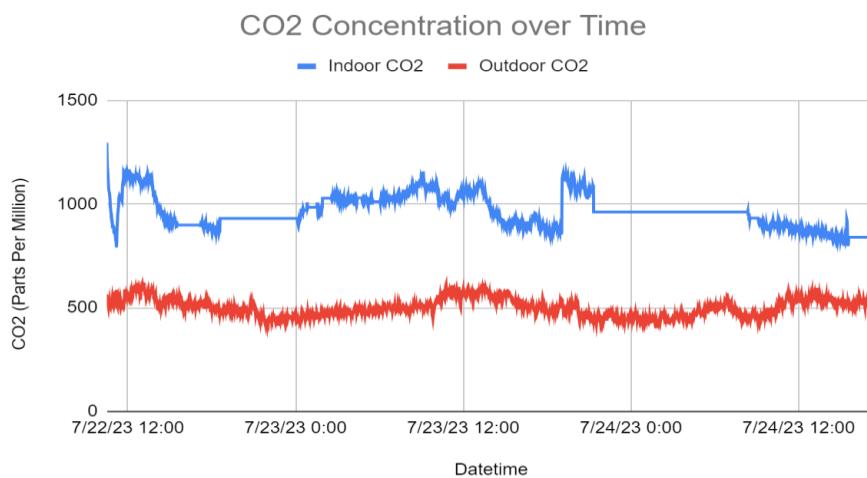
**Figure 19: PM<sub>2.5</sub> Concentration over Time Inside Dorm Room 2 and Outside**

The O<sub>3</sub> concentration inside the dorm room was significantly lower than the concentration outside. The outdoor concentration was 31.24 ppb and the indoor concentration was 9.55 ppb. This is a 69% decrease in O<sub>3</sub> concentration. (See figure 20 for a comparison of indoor and outdoor O<sub>3</sub>) Unlike the other dorm room with an air purifier, the dorm room with an open window experienced cyclical change in O<sub>3</sub> concentration according to the outdoor pattern of O<sub>3</sub> concentration.



**Figure 20: O<sub>3</sub> Concentration over Time Inside Dorm Room 2 and Outside**

The CO<sub>2</sub> concentration in dorm 2 room had fewer, shorter spikes compared to the other indoor locations tested. (See figure 21 for a comparison of outdoor and indoor concentrations of CO<sub>2</sub>) The average value of CO<sub>2</sub> in the dorm room was 963.39 ppm. However, none of the CO<sub>2</sub> peaks measured above 1200 ppm after its initial placement. This suggests that the CO<sub>2</sub> levels in this room were much more consistent than the levels in the other rooms that were tested.



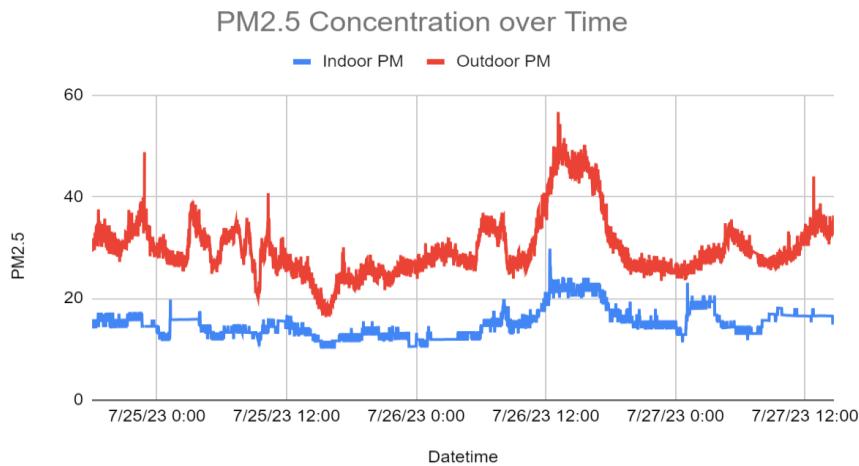
**Figure 21: CO<sub>2</sub> Concentration over Time Inside Dorm Room 2 and Outside**

#### 4. Comparison of Indoor and Outdoor Air Quality in Doherty Hall Classroom A302

Between 6:00 PM on July 24 and 2:40 PM on July 27th, the indoor sensors recorded air quality data inside Doherty Hall classroom A302, a lecture hall used for Pennsylvania Governor's School for the Sciences classes. The sensors reported a significantly lower concentration of O<sub>3</sub> and PM<sub>2.5</sub> in the classroom compared to the outdoor air. The CO<sub>2</sub> measurements showed an increase when students were in classes taking place in that classroom.

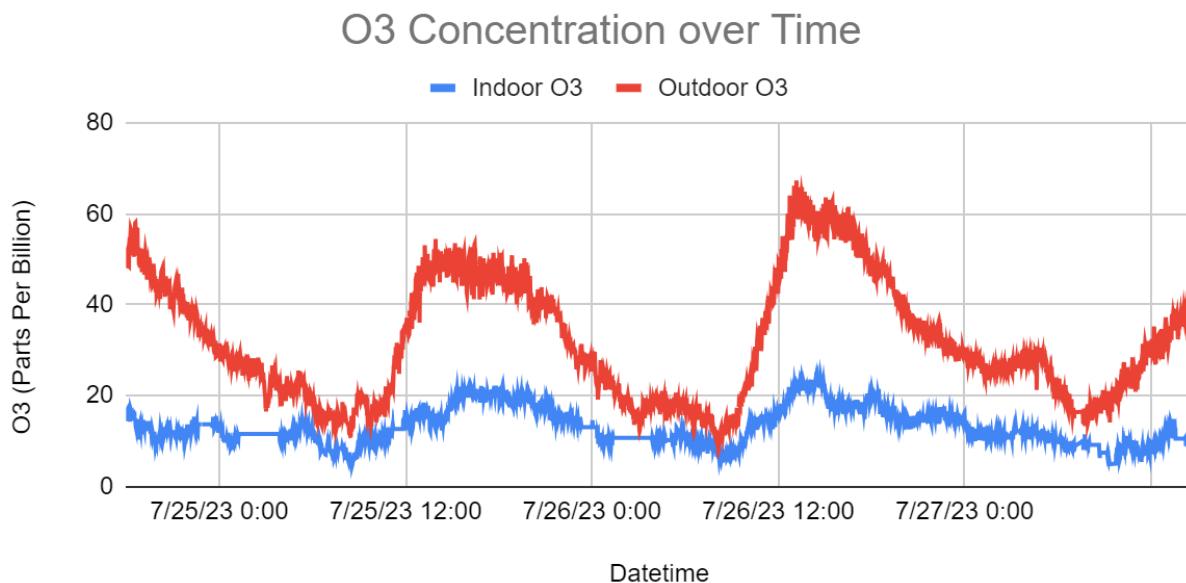
The concentration of PM<sub>2.5</sub> inside the classroom was significantly lower than the concentration outside. The average value of PM<sub>2.5</sub> in the classroom was 15.21 µg/m<sup>3</sup>, which is a 49% decrease from the outdoor

average value of 30.00  $\mu\text{g}/\text{m}^3$ . The PM<sub>2.5</sub> concentration indoors strongly mirrored the patterns of the concentration outdoors. (See figure 22 for PM<sub>2.5</sub> concentration over time) Although this is a significant decrease indoors for PM<sub>2.5</sub>, the concentration is still above the recommended exposure of 12.5  $\mu\text{g}/\text{m}^3$ .



**Figure 22: PM<sub>2.5</sub> Concentration over Time Inside Classroom and Outside**

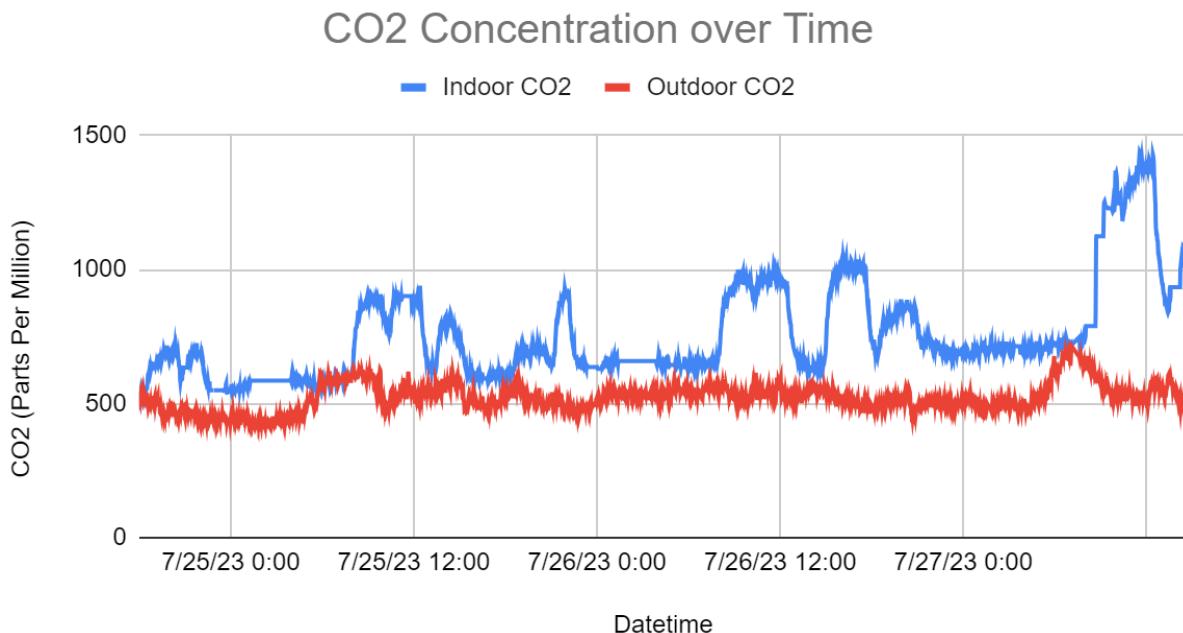
The O<sub>3</sub> concentration inside the lecture hall was significantly less than the concentration outside. The outdoor O<sub>3</sub> concentration average was 32.55 ppb and the indoor average concentration was 13.25. (See figure 23 for O<sub>3</sub> concentration over time in the classroom compared to outside) This is a 59% decrease in O<sub>3</sub> concentration. Like most other rooms tested, the classroom's O<sub>3</sub> concentration was strongly correlated to the patterns of O<sub>3</sub> concentration outdoors.



**Figure 23: O<sub>3</sub> Concentration over Time Inside Classroom and Outside**

The CO<sub>2</sub> concentration in the classroom was variable and changed depending on the presence of students. Between 8:00 AM and 12:30 PM, the concentration of CO<sub>2</sub> in the room would increase to nearly 1000 ppm most days. The concentration rose to a maximum of 1436 ppm at 11:34 AM on July 27. This is much lower

than the peaks reached in the common area even though more students occupied the classroom than the common area during their respective busiest times. (See figure 24 for CO<sub>2</sub> concentration over time in the classroom compared to the air outside) The lower concentration of CO<sub>2</sub> and shorter peaks suggests that classroom A302 has a better ventilation system compared to the common area.



**Figure 24: CO<sub>2</sub> Concentration over Time Inside Classroom and Outside**

### 5. Evaluation of Air Exchange Rate in Indoor Spaces

Air exchange rate, measured in air changes per hour, identifies the ventilation rate of a room or space. This describes the amount of time taken for a certain amount of a pollutant to diffuse out of a room. Carbon dioxide (CO<sub>2</sub>) is a great indicator for air exchange rate because it is easy to produce in large quantities, easy to stop producing, inexpensive, and safe compared to other gaseous indicators. In this experiment, air exchange rate was calculated based on the rate of CO<sub>2</sub> decrease after all occupants left a room.

$$\text{Air Exchange Rate} = A_D = \frac{1}{\Delta t} \ln\left(\frac{C_1 - C_R}{C_0 - C_R}\right) \quad (5)$$

### 6. Evaluation of Air Exchange Rate in Donner House Common Area

Using CO<sub>2</sub> data, the average air exchange rate was found to be 0.885 exchanges per hour. This was calculated using data from July 7th, 10th, and 11th from 12:00 AM until 2:00 AM. On the 7th, the air exchange rate was 0.949 exchanges per hour. On the 10th, the air exchange rate was 1.079 exchanges per hour. On the 11th, the air exchange rate was 0.626 exchanges per hour. The other days' data was not as regular because the room's occupancy decreased before all people left. This caused a CO<sub>2</sub> dip before the exchange rate could properly be measured. Although the CO<sub>2</sub> levels were decreasing, there were people in the room, so it was being filled with new CO<sub>2</sub> that would make the air exchange data inaccurate.

The ASHRAE recommends 0.35 air changes per hour in residential buildings such as houses.<sup>46</sup> The common area is far more crowded than a residential building would be, so this recommended air exchange rate would not be safe in such a setting. Several states have laws firmly recommending classrooms to have

at least 2, or in some cases, at least 6, air changes per hour.<sup>47</sup> This means that the common area in Donner House has a potentially unsafe air exchange rate. If a large group of people, such as the students who used it to study during Pennsylvania Governor's School for the Sciences, remained in that room for several hours, it may result in an irritating or even dangerous concentration of CO<sub>2</sub>.

### 7. Evaluation of Air Exchange Rate in Donner House Dorm Rooms

The air exchange rate of the dorm room with the air conditioner and air purifier was calculated to be 1.535 air changes per minute between 7:30 AM and 8:30 AM on July 20th. However, this number may be artificially high due to the door being opened in the morning. Using a two-hour average from 7:30 to 9:30 gives an air exchange rate of 0.768 air changes per hour. This number is relatively low for such a small room. Whereas this low air exchange rate may be safe for two occupants, it poses a potential hazard if several other people inhabit the room for an extended period of time.

The air exchange rate of the dorm room with an open window cannot accurately be measured given the data. That room does not experience any major spikes in CO<sub>2</sub> overnight, so it just fluctuates on a smaller scale. This suggests that the air exchange rate overpowers the rate of CO<sub>2</sub> production. To measure the air exchange rate accurately would require an excess of some gas or substance to slowly diffuse out of a room. If that rate of diffusion is greater than the rate of gas production, a considerable excess cannot be accumulated. Based on CO<sub>2</sub> alone, the dorm room with an open window is safer than the dorm room with air conditioning and an air purifier.

### 8. Evaluation of Air Exchange Rate in Doherty Hall Classroom A302

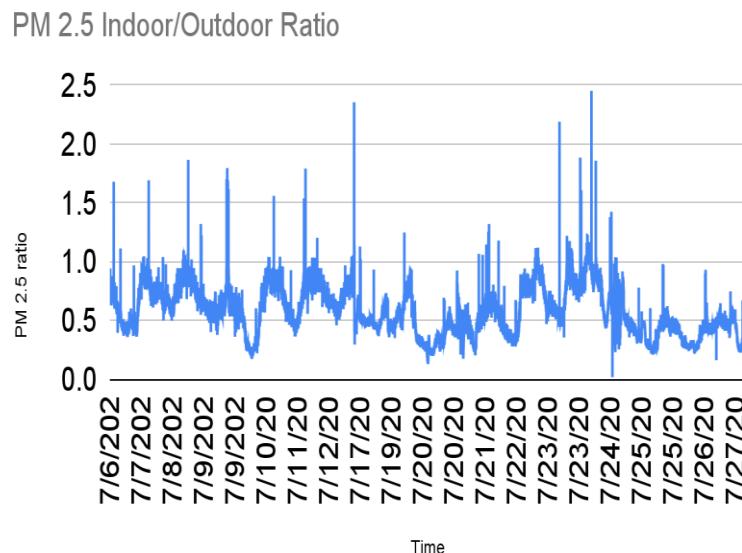
Using data from 12:30 PM to 1:30 PM on July 25th, the calculated air exchange rate in Doherty Hall classroom A302 is 5.949 air changes per hour. This ventilation rate is nearly identical to the recommended classroom ventilation rate of 6 air changes per hour. Using data from 12:30 PM until 1:30 PM on July 26th, the calculated air exchange rate is 4.928 air changes per hour. This value is slightly further from the ideal recommended classroom ventilation rate of 6 air changes per hour. Still, it is in the range of safe values and does not pose any risk to health.

## C. Indoor/Outdoor Ratios

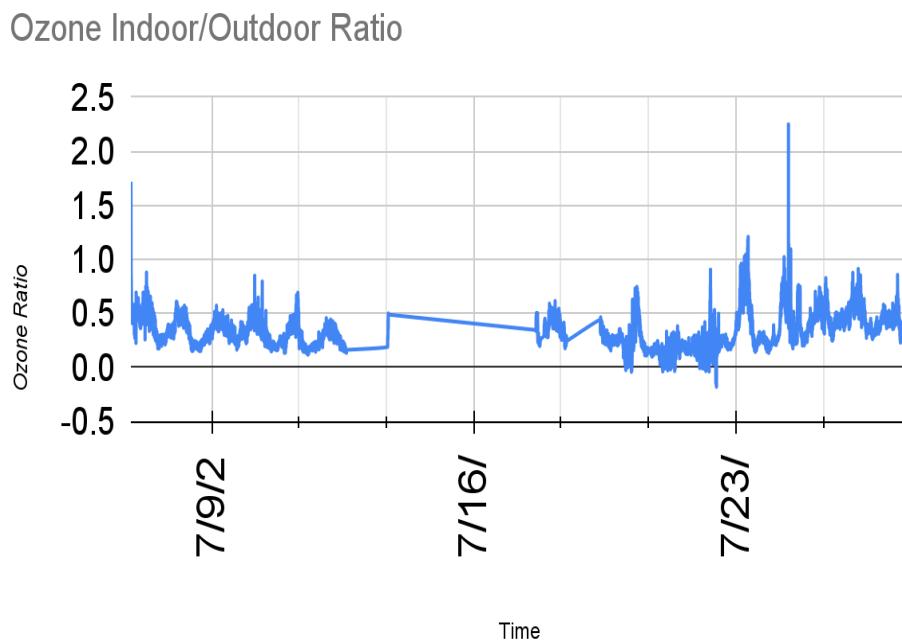
Indoor/outdoor ratios are used to compare the results of the experiment. These ratios are found with O<sub>3</sub> and PM<sub>2.5</sub> measurements collected by the indoor and outdoor sensors. To find the indoor/outdoor ratios Equation 5 must be used.

$$\frac{\text{Indoor Pollutant Concentration}}{\text{Outdoor Pollutant Concentration}} \quad (6)$$

In Equation 6 the indoor pollutant concentration of a given time is divided by the outdoor pollutant concentration at the same given time on the same day. This shows how much more of the pollutant concentration there was outdoors compared to indoors. When looking at the ratios in figures 25 and 26 many interesting conclusions can be drawn. In the O<sub>3</sub> indoor/outdoor ratio of Figure 26, from 7/19-7/22 there are multiple times where there are points at approximately 0. This means that there was little to no O<sub>3</sub> in dorm 1. This is most likely because dorm 1 had an air purifier. It can also be noted that there is a daily trend for both PM<sub>2.5</sub> and O<sub>3</sub> which is shown by all of the peaks and valleys. There is also a gap in the data due to the indoor sensor being offline for that time.



**Figure 25: PM<sub>2.5</sub> Indoor/Outdoor Ratio Over the 20 Days Recorded**

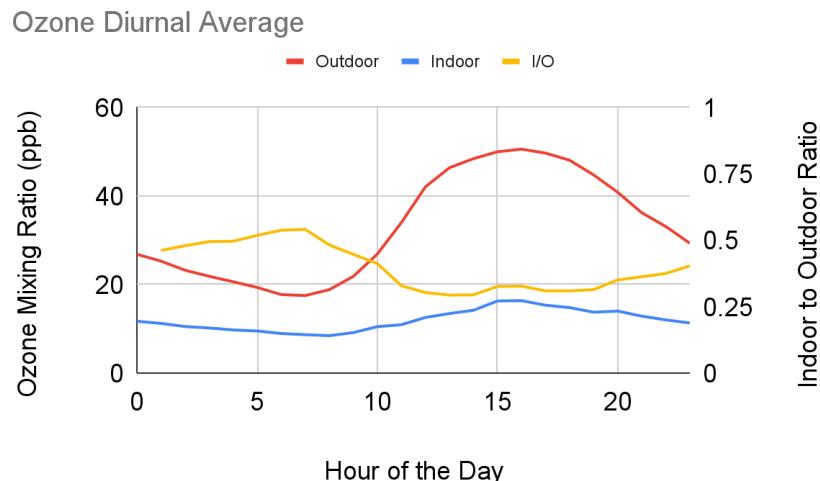


**Figure 26: O<sub>3</sub> Indoor/Outdoor Ratio Over the 20 Days Recorded**

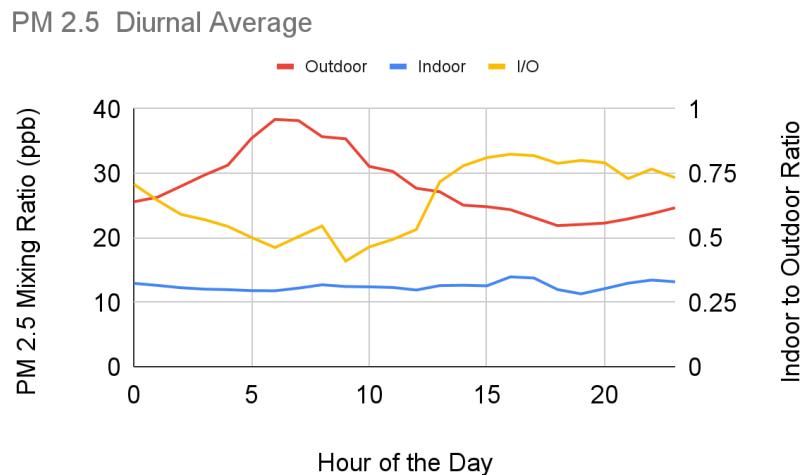
## D. Diurnal Patterns

Diurnal patterns show the pattern that the pollutant concentrations follow throughout the day. The diurnal patterns are found by calculating the hourly average of the pollutant concentration. O<sub>3</sub> outdoors follows a

diurnal pattern that shows its highest reading at approximately the 16th hour of the day and its lowest point is at approximately the 7th hour of the day as seen in Figure 27. The O<sub>3</sub> is higher in the afternoon due to photochemistry. Photochemistry needs to absorb light to occur, so it is more present in the afternoon when the sun is up.<sup>48</sup> The O<sub>3</sub> indoors follows the diurnal pattern in the same way only the highs and lows are much smaller. The diurnal patterns for PM<sub>2.5</sub> outdoors, show that it is highest at approximately the 6th hour of the day and lowest at approximately the 22nd hour of the day as seen in Figure 28. This is due to the confinement of the mixing layer in the morning.<sup>49</sup> Indoors, the PM<sub>2.5</sub> concentration seemed to remain relatively constant throughout the day. This is most likely due to ventilation and air filtration indoors.



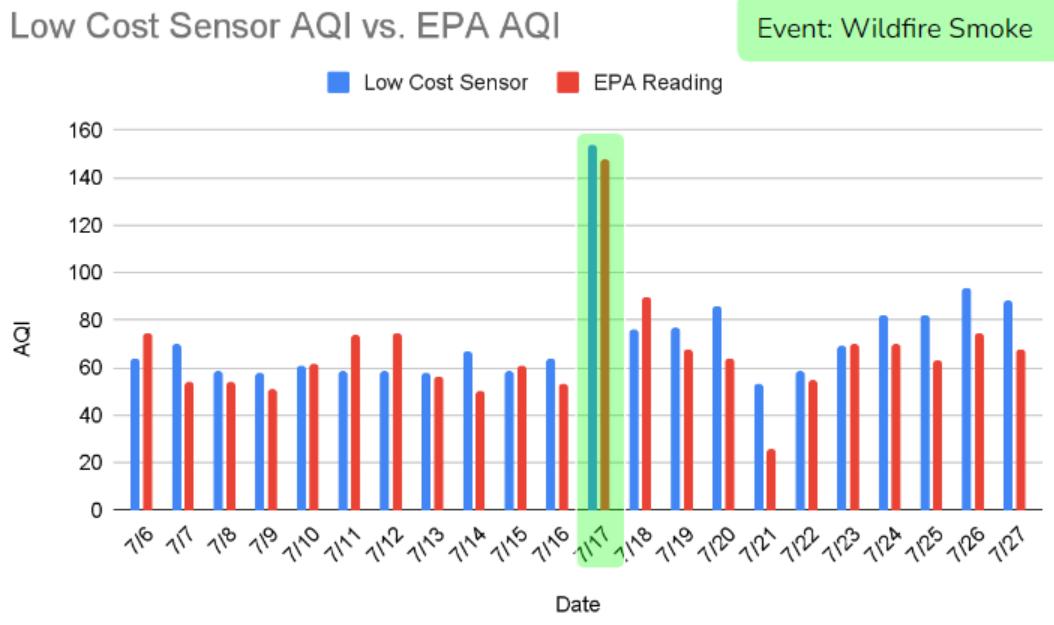
**Figure 27: O<sub>3</sub> Average Diurnal Patterns Over the 20 Days Recorded**



**Figure 28: PM<sub>2.5</sub> Average Diurnal Patterns Over the 20 Days Recorded**

## E. Indoor and Outdoor AQI Comparisons

In order to further determine whether the low-cost sensors were able to produce accurate data in regard to the air quality students were exposed to during the PGSS program, the AQI values calculated by the low-cost sensor were compared to those released by the EPA based on the data gathered by regulatory monitoring station sensors. Figure 29 illustrates a comparison between the outdoor AQI calculated by the low-cost sensor and the AQI calculated by the regulatory monitoring station.



**Figure 29: Comparison of Low-cost Sensor AQI vs. EPA AQI**

It can be observed in Figure 29 that both the low-cost sensor AQI and the EPA AQI follow roughly the same trends. This is with a few exceptions, however. Most discrepancies between the data can be attributed to local events. This means that there was likely an event that occurred on campus that impacted the nearby (low-cost) sensors, but not the EPA's sensors that are located elsewhere. Some possible examples of such events include construction and grass cutting, which occurred often on campus throughout the duration of PGSS.

On one day in particular, July 21st, the low-cost sensor measured a significantly higher AQI than the EPA sensor did. The low-cost sensor reported an AQI of 53, while the EPA reported an AQI of 26. When an analysis of the data from the 21st was conducted, it appeared as though the PM<sub>2.5</sub> concentrations peaked during the morning hours (around 8:00am - 9:30am). It is possible that some yard work was done around the sensor at that time, but there is no guarantee. The same possibility of local events could be the cause for the EPA AQI measurements that are higher than the low-cost sensor measurements. A very large portion of the readings given by an air sensor depend on the events that occur in the area directly surrounding it.

One event, occurring on a much larger scale, was able to have a substantial impact on both the low-cost AQI and the EPA AQI. This event was caused by the wildfires occurring in Canada over the duration of PGSS. The fires occurring throughout Canadian forests produced extreme amounts of smoke. This smoke was carried south into the United States by pressure systems and wind. The smoke being carried into the

US caused an extreme peak in PM<sub>2.5</sub> concentration measurements. As seen in the graph, one of these peaks occurred on July 17th. The PM<sub>2.5</sub> measurements taken on this day are substantially higher than the surrounding days. This directly correlates to an increase in wildfire smoke seen in Pittsburgh. The EPA sensor measured concentrations that correlate to an AQI of 148, which falls in the category of being unhealthy for sensitive groups. The low-cost sensor, however, measured concentrations correlating to an AQI value of 154, which is quite like the EPA value, but falls in a different category: unhealthy for all people. This difference could mean that the smoke was, on average, worse on the CMU campus than throughout the rest of Pittsburgh.

Because PGSS students spend so much of their time indoors during the program, it is important for the data to also compare indoor air quality to outdoor air quality. This comparison can demonstrate whether the air quality indoors is better than the air quality outdoors. Figure 30 demonstrates such differences. The dates are color coded depending on the location of the indoor sensor at the time the measurements were being taken. Figure 30 is a key that translates colors into locations.

### Outdoor Sensor AQI vs. Indoor Sensor AQI

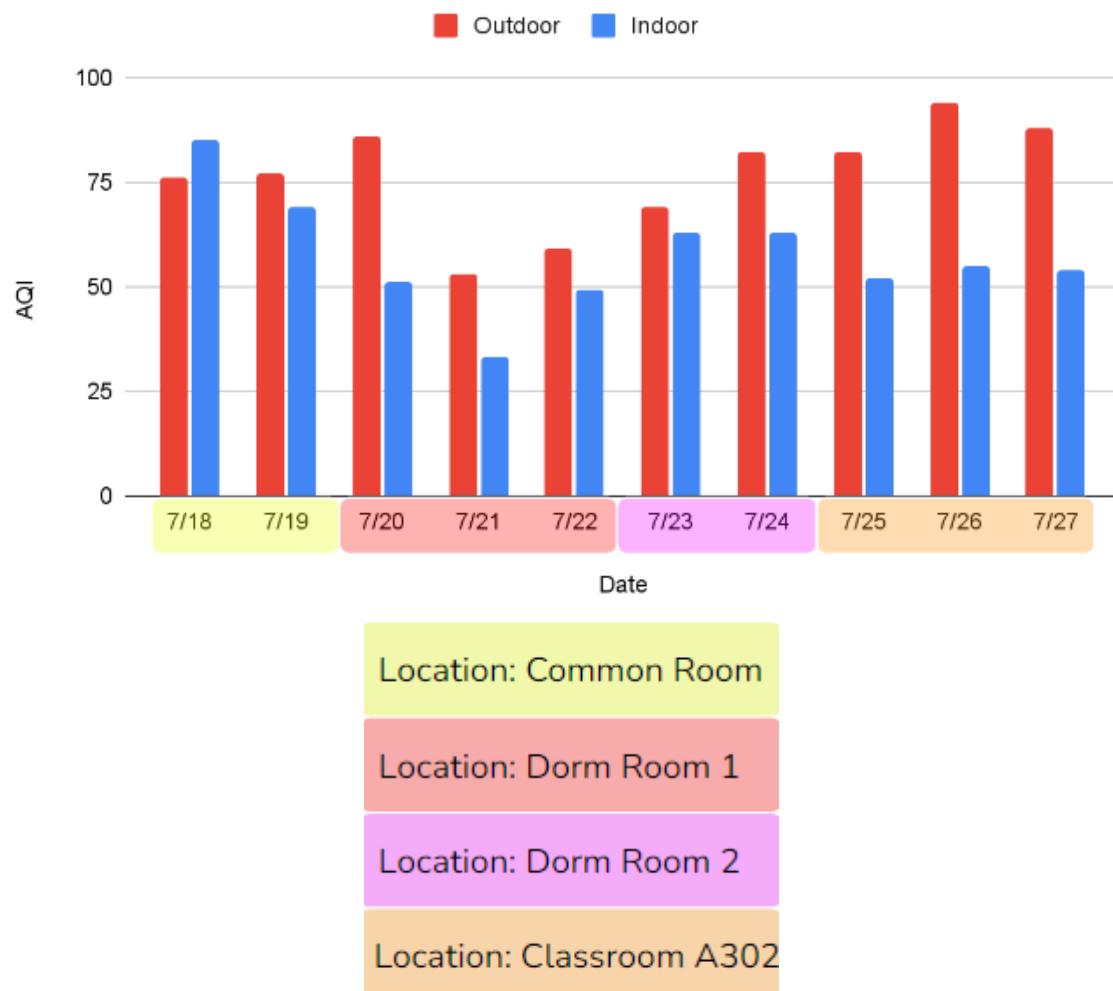


Figure 30: Air Quality index categorized by the indoor sampling location.

Based on the general peaks of the data, it can be concluded that the air quality indoors is better than the air quality outdoors. Only one out of the 10 peaks representing indoor air quality is higher than the peak representing the outdoor air quality. The difference between the indoor and outdoor air quality differs depending on the exact location of the indoor sensor. By subtracting the AQI of the indoor sensor from the AQI of outdoor sensor, an AQI difference was calculated, then averaged for each location of the sensor. The average differences are represented in Table 3.

**Table 3: Average AQI Difference Depending on Location of Indoor Sensor**

Location	Average AQI Difference
Common Room	34
Dorm Room 1	22
Dorm Room 2	12.5
Classroom A302	34

The higher the difference between the AQI values, the better the indoor air quality is compared to the outdoor air quality. Both the common room and classroom A302 have average differences of 34, meaning that the air quality in both of those rooms is substantially better than the air quality measured outside. Dorm room 1, a room with an air conditioner and air purifier falls beneath those differences at a difference of 22. This was a slightly unexpected result, as the air purifier was assumed to be an effective method to decrease PM<sub>2.5</sub> concentrations within a room. The calculated average AQI difference for dorm room 2, however, was as expected, as dorm room 2 had its windows open. This meant that air from outdoors was easily able to flow into the room, causing the indoor air to consist of many of the same concentrations of particles that the outdoor air did. In general, it can be concluded from this data that, depending on the conditions of the room, the air quality was very often better inside than it was outside throughout the duration of the PGSS program.

## IV. Discussion

### A. Health Implications

The concentrations found in the sensor data are associated with many minor health risks. The outdoor PM<sub>2.5</sub> and indoor CO<sub>2</sub> values present a potential hazard to sensitive groups. The O<sub>3</sub> concentration never reached unsafe levels inside or outside.

The outdoor PM<sub>2.5</sub> value could cause several minor negative health effects. Short-term exposure to PM<sub>2.5</sub> concentrations above 75, such as on July 17th, can cause coughing, shortness of breath, headaches, and other mild symptoms. These symptoms would be much worse in people with respiratory conditions such as asthma. Additionally, the everyday exposure to PM<sub>2.5</sub> measured throughout the experiment could eventually lead to worsening of respiratory conditions or respiratory inflammation in the long term. While the PM<sub>2.5</sub> concentration indoors was mostly safe, it did reach above 50 on July 17th. This could cause negative health effects in sensitive individuals. Even when the indoor PM<sub>2.5</sub> concentrations were safe, they were on average above the 12.5 recommended daily average in all locations other than the dorm room with air conditioning and an air purifier.

The CO<sub>2</sub> concentrations indoors were alarming. In the common area, CO<sub>2</sub> buildup significantly exceeded the sensors' maximum range. This is a health risk. Exposure to CO<sub>2</sub> concentrations over 2000 ppm is known to cause nausea, headaches, drowsiness, and dizziness. This was especially bad on the night of July 18th, with an estimated CO<sub>2</sub> concentration of roughly 3500 ppm. This value had to be estimated based on the trajectory of the graph of CO<sub>2</sub> since the CO<sub>2</sub> concentration significantly exceeded the sensor's maximum value. The common area's air exchange rate of 0.885 air changes per hour and its high occupancy create a potential risk of critical CO<sub>2</sub> buildup. CO<sub>2</sub> is produced far faster than it can be diffused out of the room.

## B. Summary

The summary of the data shows that the best air quality is found indoors. Specifically in a room with an air purification system or efficient filtration system. This conclusion can be made using the data that compares the outdoor versus indoor AQI graph and the indoor/outdoor ratio graphs. Furthermore, in the AQI graph, it can be observed that there is a more significant difference in indoor and outdoor air quality on the days when the sensor was in classroom A302, likely because of the building air purification system. Overall, from the data collected one can conclude that being indoors will provide protection from the typically higher PM<sub>2.5</sub> concentrations outside and that while inside you can risk being exposed to a higher concentration of CO<sub>2</sub>.

It can be concluded that to live a healthy lifestyle it is important to stay mindful of large groups indoors, stay indoors during air quality events, and invest in an air purifier. Most importantly, it can be concluded that low-cost could indeed accurately measure air quality specific to the air PGSS students were exposed to during their time at CMU.

## References

<sup>1</sup> Centers for Disease Control and Prevention. (2022, November 21). Air Pollutants. Centers for Disease Control and Prevention. <https://www.cdc.gov/air/pollutants.htm>

<sup>2</sup> Environmental Protection Agency. (2023, July 11). Particulate Matter (PM) Basics. EPA. <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>

<sup>3</sup> Kumar, D., & Deepak, K. (2018). Size of Particle - Dust Control. Size of Particle - an overview | ScienceDirect Topics. <https://www.sciencedirect.com/topics/engineering/size-of-particle#:~:text=Particles%20of%20sizes%20in%20the,near%20the%20source%20of%20formation>

<sup>4</sup> NASA. (2023, August 2). Carbon Dioxide Concentration. NASA. <https://climate.nasa.gov/vital-signs/carbon-dioxide/>

<sup>5</sup> Lindsey, R. (2023, May 12). Climate Change: Atmospheric Carbon Dioxide. NOAA Climate.gov. <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide#:~:text=Carbon%20dioxide%20is%20Earth's%20most.including%20back%20toward%20Earth's%20surface>

<sup>6</sup> FSIS Environmental Safety and Health Group. (2020, August). Carbon Dioxide Health Hazard Information Sheet.

<sup>7</sup> Iowa Department of Natural Resources. (2023). Environmental Protection. Iowa Department of Natural Resources. <https://www.iowadnr.gov/Environmental-Protection/Air-Quality/Air-Pollutants/Effects->

Ozone#:~:text=Ozone%20damages%20vegetation%20and%20ecosystems,process%20and%20release%20as%20oxygen

<sup>8</sup> Department of Environmental Protection. (2023). PA Mining History. Coal Mining In Pennsylvania PA Mining History. <https://www.dep.pa.gov/Business/Land/Mining/Pages/PA-Mining-History.aspx>

<sup>9</sup> Vintage Menu Art UK. (2023). Pittsburgh, Meda Logan Poem 1907. Vintage Menu Art . <https://www.vintagemenuart.co.uk/products/pittsburgh-med-a-logan-poem-1907>

<sup>10</sup> Martin, R. (2023, April 18). Carnegie Steel Company. Encyclopædia Britannica. <https://www.britannica.com/topic/Carnegie-Steel-Company>

<sup>11</sup> worldsteel Association. (2022, September 23). Climate Change Policy Paper. Climate change and the production of iron and steel. <https://worldsteel.org/publications/policy-papers/climate-change-policy-paper/>

<sup>12</sup> Tarr, J. (2003). Ensuring Environmental Health in Postindustrial Cities: Workshop Summary. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/books>

<sup>13</sup> Byrnes, M. (2012, June 5). What Pittsburgh Looked Like When It Decided It Had a Pollution Problem. Bloomberg.com. <https://www.bloomberg.com/news/articles/2012-06-05/what-pittsburgh-looked-like-when-it-decided-it-had-a-pollution-problem>

<sup>14</sup> University of Pittsburgh. (2023). Historic Pittsburgh. <https://historicpittsburgh.org/>

<sup>15</sup> Luconi, S. (1999, Autumn). The Enforcement of the 1941 Smoke-Control Ordinance and Italian Americans in Pittsburgh - JSTOR. JSTOR. <https://www.jstor.org/stable/27774217>

<sup>16</sup> Tarr, J. (2003). Ensuring Environmental Health in Postindustrial Cities: Workshop Summary. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/books>

<sup>17</sup> Jacobs, E. T., Burgess, J. L., & Abbott, M. B. (2018, April). The Donora Smog Revisited: 70 years after the event that inspired the Clean Air Act. American journal of public health. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5922205/>

<sup>18</sup> Environmental Protection Agency. (2022, November 28). Evolution of the Clean Air Act. EPA. <https://www.epa.gov/clean-air-act-overview/evolution-clean-air-act>

<sup>19</sup> Minnesota Pollution Control Agency. (2023). Understanding the air quality index (AQI). [https://www.pca.state.mn.us/air-water-land-climate/understanding-the-air-quality-index-aqi#:~:text=The%20Air%20Quality%20Index%20\(AQI\)%20was%20developed%20by%20the%20U.S](https://www.pca.state.mn.us/air-water-land-climate/understanding-the-air-quality-index-aqi#:~:text=The%20Air%20Quality%20Index%20(AQI)%20was%20developed%20by%20the%20U.S)

<sup>20</sup> AirNow.gov, U.S. EPA. (203AD). Using Air Quality index. Using Air Quality Index | AirNow.gov. <https://www.airnow.gov/aqi/aqi-basics/using-air-quality-index/#:~:text=The%20AQI%20as%20we%20know,health%2Dbased%20air%20quality%20standards>

<sup>21</sup> Phox, J. (2023, August 1). *Summit held to discuss Pittsburgh's ongoing air quality issues and high asthma rates*. Pittsburgh City Paper. <https://www.pghcitypaper.com/news/summit-held-to-discuss-pittsburghs-ongoing-air-quality-issues-and-high-asthma-rates-20520435#:~:text=Allegheny%20County%20has%20a%20high,than%20the%208%25%20national%20average>

<sup>22</sup> Asthma & Allergy Foundation of America. (2022, November 30). Asthma Capitals. Asthma & Allergy Foundation of America. <https://aafa.org/asthma-allergy-research/our-research/asthma-capitals/>

<sup>23</sup> Czachor, E. M. (2023, July 19). Canadian wildfire maps show where fires continue to burn across Quebec, Ontario and other provinces. CBS News. <https://www.cbsnews.com/news/map-canadian-wildfires-2023-where-are-the-fires-ontario-quebec/>

<sup>24</sup> Canada, N. R. (2023). Interactive map. Canadian Wildland Fire Information System. <https://cwfis.cfs.nrcan.gc.ca/interactive-map?zoom=1%C2%A2er&month=7&day=15&year=2023#iMap>

<sup>25</sup> Sokhi, R. S., Moussiopoulos, N., Baklanov, A., Bartzis, J., Coll, I., Finardi, S., Friedrich, R., Geels, C., Grönholm, T., Halenka, T., Ketzel, M., Maragkidou, A., Matthias, V., Moldanova, J., Ntziachristos, L., Schäfer, K., Suppan, P., Tsegas, G., Carmichael, G., Franco, V., Hanna, S., Jalkanen, J.-P., Velders, G. J. M., and Kukkonen, J.: Advances in air quality research – current and emerging challenges, *Atmos. Chem. Phys.*, 22, 4615–4703, <https://doi.org/10.5194/acp-22-4615-2022>, 2022.

<sup>26</sup> Victoria, E. P. A. (2021, August 23). PM2.5 particles in the Air. EPA Victoria. <https://www.epa.vic.gov.au/for-community/environmental-information/air-quality/pm25-particles-in-the-air>

<sup>27</sup> Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis, Martin Adam, Tamara Schikowski, Anne Elie Carsin, Yutong Cai, Benedicte Jacquemin, Margaux Sanchez, Andrea Vierkötter, Alessandro Marcon, Dirk Keidel, Dorothee Sugiri, Zaina Al Kanani, Rachel Nadif, Valérie Siroux, Rebecca Hardy, Diana Kuh, Thierry Rochat, Pierre-Olivier Bridevaux, Marloes Eeftens, Ming-Yi Tsai, Simona Villani, Harish Chandra Phuleria, Matthias Birk, Josef Cyrys, Marta Cirach, Audrey de Nazelle, Mark J. Nieuwenhuijsen, Bertil Forsberg, Kees de Hoogh, Christophe Declercq, Roberto Bono, Pavilio Piccioni, Ulrich Quass, Joachim Heinrich, Deborah Jarvis, Isabelle Pin, Rob Beelen, Gerard Hoek, Bert Brunekreef, Christian Schindler, Jordi Sunyer, Ursula Krämer, Francine Kauffmann, Anna L. Hansell, Nino Künzli, Nicole Probst-Hensch, European Respiratory Journal Jan 2015, 45 (1) 38-50; DOI: 10.1183/09031936.00130014

<sup>28</sup> Chen J, Rodopoulou S, de Hoogh K, Strak M, Andersen Z, Atkinson R, Bauwelinck M, Bellander T, Brandt J, Cesaroni G, Concin H, Fecht D, Forastiere F, Gulliver J, Hertel O, Hoffmann B, Hvidtfeldt U, Janssen N, Jöckel K, Jørgensen J, Katsouyanni K, Ketzel M, Klompmaker J, Lager A, Leander K, Liu S, Ljungman P, MacDonald C, Magnusson P, Mehta A, Nagel G, Oftedal B, Pershagen G, Peters A, Raaschou-Nielsen O, Renzi M, Rizzato D, Samoli E, van der Schouw Y, Schramm S, Schwarze P, Sigsgaard T, Sørensen M, Stafoggia M, Tjønneland A, Vienneau D, Weinmayr G, Wolf K, Brunekreef B and Hoek G (2021) Long-Term Exposure to Fine Particle Elemental Components and Natural and Cause-Specific Mortality—a Pooled Analysis of Eight European Cohorts within the ELAPSE Project, *Environmental Health Perspectives*, 129:4, Online publication date: 1-Apr-2021.

<sup>29</sup> Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z. J., Badaloni, C., Beelen, R., Caracciolo, B., Faire, U. de, Erbel, R., Eriksen, K. T., Fratiglioni, L., Galassi, C., Hampel, R., Heier, M., Hennig, F., Hilding, A., Hoffmann, B., Houthuijs, D., Jöckel, K.-H., ... Peters, A. (2014, January 21). Long term exposure to ambient air pollution and incidence of acute coronary events: Prospective cohort study and meta-analysis in 11 European cohorts from the escape project. *The BMJ*. <https://www.bmjjournals.org/content/348/bmj.f7412>

<sup>30</sup> Occupational Safety and Health Administration. (n.d.). OZONE. United States Department of Labor. <https://www.osha.gov/chemicaldata/9>

<sup>31</sup> MIT. (n.d.). What is the ideal level of carbon dioxide in the atmosphere for human life?. MIT Climate Portal. <https://climate.mit.edu/ask-mit/what-ideal-level-carbon-dioxide-atmosphere-human-life>

<sup>32</sup> Occupational Safety and Health Administration. (n.d.-a). CARBON DIOXIDE. United States Department of Labor. <https://www.osha.gov/chemicaldata/183>

<sup>33</sup> Carbon dioxide. Wisconsin Department of Health Services. (2023, March 29). <https://www.dhs.wisconsin.gov/chemical/carbon dioxide.htm>

<sup>34</sup> Beta attenuation technology for particulate matter measurement. Thermo Fisher Scientific - US. (n.d.). <https://www.thermofisher.com/us/en/home/industrial/environmental/environmental-learning-center/air-quality-analysis-information/beta-attenuation-technology-particulate-matter-measurement.html>

<sup>35</sup> United States Environmental Protection Agency. (n.d.-c). EPA scientists develop and evaluate Federal Reference & Equivalent Methods for measuring key air pollutants. EPA. <https://www.epa.gov/air-research/epa-scientists-develop-and-evaluate-federal-reference-equivalent-methods-measuring-key>

<sup>36</sup> Chris Hubbuch | Wisconsin State Journal. (2022, November 27). Madison to use low-cost sensors to measure neighborhood air quality. Wisconsin State Journal. [https://madison.com/news/local/environment/madison-to-use-low-cost-sensors-to-measure-neighborhood-air-quality/article\\_558f2731-5094-5b7f-b28a-a40a0ca08b3e.html](https://madison.com/news/local/environment/madison-to-use-low-cost-sensors-to-measure-neighborhood-air-quality/article_558f2731-5094-5b7f-b28a-a40a0ca08b3e.html)

<sup>37</sup> COMMUNITY AIR QUALITY MONITORING WITH THE RAMP SENSOR PACKAGE. R Subramanian. (n.d.). <https://www.rsubramanian.com/ramp-sensors.html>

<sup>38</sup> Tanzer R, Malings C, Hauryliuk A, Subramanian R, Presto AA. Demonstration of a Low-Cost Multi-Pollutant Network to Quantify Intra-Urban Spatial Variations in Air Pollutant Source Impacts and to Evaluate Environmental Justice. Int J Environ Res Public Health. 2019 Jul 15;16(14):2523. doi: 10.3390/ijerph16142523. PMID: 31311099; PMCID: PMC6678618.

<sup>39</sup> Freeman, G. (2023, February 22). How do electrochemical gas sensors work?: Nova gas analyzers. Nova Gas. <https://www.nova-gas.com/how-do-electrochemical-gas-sensors-work/>

<sup>40</sup> Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., Presto, A. A., and R. Subramanian: Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring, Atmos. Meas. Tech., 12, 903–920, <https://doi.org/10.5194/amt-12-903-2019>, 2019.

<sup>41</sup> U.S. Environmental Protection Agency Office of Quality Planning and Standards. (2018, September). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index(AQI).ResearchTrianglePark,NC.<https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf>

<sup>42</sup> Carl Malings, Rebecca Tanzer, Aliaksei Hauryliuk, Provat K. Saha, Allen L. Robinson, Albert A. Presto & R Subramanian (2020) Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation, Aerosol Science and Technology, 54:2, 160-174, DOI: 10.1080/02786826.2019.1623863

<sup>43</sup> Environmental Protection Agency. (n.d.). Air Data Basic Information. EPA. [https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information#:~:text=The%20AQI%20\(Air%20Quality%20Index\)&text=The%20AQI%20is%20an%20index,runs%20from%200%20to%20500](https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information#:~:text=The%20AQI%20(Air%20Quality%20Index)&text=The%20AQI%20is%20an%20index,runs%20from%200%20to%20500)

<sup>44</sup> AirNow.gov, U.S. EPA. (n.d.). Air Quality Index (AQI) Basics. AQI Basics | AirNow.gov. <https://www.airnow.gov/aqi/aqi-basics/>

<sup>45</sup> Minnesota Pollution Control Agency. (n.d.). Understanding the air quality index (AQI). Understanding the Air Quality Index (AQI). <https://www.pca.state.mn.us/air-water-land-climate/understanding-the-air-quality-index-aqi>

<sup>46</sup> Environmental Protection Agency. (n.d.-b). How much ventilation do I need in my home to improve indoor air quality?. EPA. [https://www.epa.gov/indoor-air-quality-iaq/how-much-ventilation-do-i-need-my-home-improve-indoor-air-quality#:~:text=ASHRAE%20\(formerly%20called%20the%20American,air%20per%20minute%20\(cfm\)%20per](https://www.epa.gov/indoor-air-quality-iaq/how-much-ventilation-do-i-need-my-home-improve-indoor-air-quality#:~:text=ASHRAE%20(formerly%20called%20the%20American,air%20per%20minute%20(cfm)%20per)

<sup>47</sup> Minimum Air Changes Per Hour (ACH) recommended for reducing COVID across 50 states. Patient knowhow. (n.d.). <https://www.patientknowhow.com/ach.html>

<sup>48</sup> Bonfield HE, Knauber T, Lévesque F, Moschetta EG, Susanne F, Edwards LJ. Photons as a 21st century reagent. Nat Commun. 2020 Feb 6;11(1):804. doi: 10.1038/s41467-019-13988-4. PMID: 32029723; PMCID: PMC7004975.

<sup>49</sup> Du, Q., Zhao, C., Zhang, M., Dong, X., Chen, Y., Liu, Z., Hu, Z., Zhang, Q., Li, Y., Yuan, R., and Miao, S.: Modeling diurnal variation of surface PM<sub>2.5</sub> concentrations over East China with WRF-Chem: impacts from boundary-layer mixing and anthropogenic emission, Atmos. Chem. Phys., 20, 2839–2863, <https://doi.org/10.5194/acp-20-2839-2020>, 2020.



**COMPUTER SCIENCE  
TEAM PROJECTS**



# Journey Through the Star: A Minimax Exploration of Chinese Checkers

Damien Busche, Yanwei Liu, Helen Mao, Olina Mukherjee, Ben Nguyen, Parth Parikh

## Abstract

Simple yet efficient, the minimax algorithm can be adopted to intelligently play turn-based games. Using the language of Python, this paper outlines a minimax-based approach to intelligently play a simplified two-player version of the game of Chinese Checkers. In addition to the overall minimax algorithm, this paper outlines possible additions to the algorithm in the form of the endgame routine and Alpha-Beta Pruning. Both of these additions to the Chinese Checkers game proved to be extremely effective at reducing runtime and increasing the win rate of the computer. Further, this research demonstrated the efficacy of incorporating aspects of board evaluation such as standard deviation and encouraging pieces to move toward the center. Interestingly, odd numbers of layers of search depth proved to be anomalies in the data collected as the algorithm attempted defensive play and was thus unable to determine a winning move.<sup>1</sup>

## I. Introduction

### A. Chinese Checkers

#### 1. History

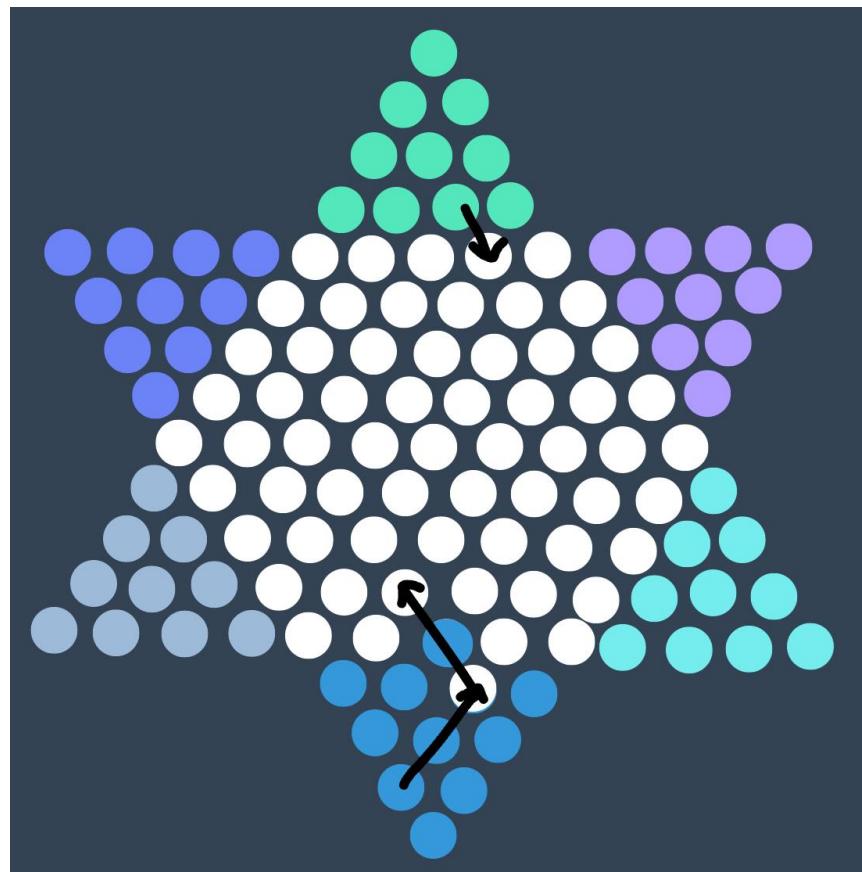
Contrary to the game's name, Chinese Checkers originated in Germany around 1892. The turn-based game was inspired by a similar game called Halma which was invented in the United States and adapted in Germany. One key difference between the games lies in their respective board shapes: Halma uses a square board with 256 holes while Chinese checkers uses a star-shaped board with 121 holes. Additionally, unlike Halma, which allows two to four players, Chinese checkers can accommodate up to six players. The first version of the new game was released under the name Stern-Halma, which literally means Star-Halma, but then Pressman & Co., an American company, purchased the game in 1928 and changed the name to "Hop Ching Checkers". Shortly after, Hop Ching Checkers became "Chinese Checkers" due to marketing purposes, and it remains known by this name today.<sup>2</sup> Today, Chinese Checkers is a common leisure activity played in East Asia and parts of Europe.<sup>3</sup>

#### 2. Gameplay

In the game, each player has a total of 10 pieces that start in their triangular home space. The objective is to then move the pieces to the opposite home space such that two conditions are met: all holes in the opposite triangle are filled and at least one of the pieces occupying the holes is from the player. During gameplay, as the category of turn-based game suggests, the players take turns moving one of their pieces. A player can only move one piece each turn, but depending on the location of the piece, they can make as many consecutive jumps as they want. There are two types of legal moves in Chinese Checkers.

**Single-Space Moves:** Single-space moves can be made by selecting a piece of choice and moving it one space diagonally to any unoccupied adjacent hole. After a single-space move is made, the player's turn ends. Figure 1 demonstrates such a move from the green player's perspective.

**Jump Moves:** A jump is a move made when a piece “jumps” over an adjacent piece. A single jump is defined as going in a linear path over one adjacent peg and ultimately results in the piece reaching a hole two spaces away from the starting position. Once a piece has initiated a jump move, it can legally keep making jump moves. Once the player decides to stop jumping, their turn comes to an end. Figure 1 demonstrates an example of such jumping from the blue player’s perspective.



**Figure 1: Jump Move in Chinese Checkers**

**Strategy:** There are multiple strategies that players must consider to select the most optimal move. One of the most common game strategies is to keep the pieces close to each other in a chain formation as they advance in order to secure the most out of a jump. With this method, the player can use the jumping chain to move multiple pegs across the board in a single move.

## B. Minimax

The minimax algorithm, also known as Min-Max, is a method of decision-making that implements a decision tree with all of the possible outcomes of the game. From these moves, the algorithm analyzes all of the plays and chooses the most optimal move in the current game state.<sup>4</sup> This minimax algorithm has been used in many decision-making applications. For instance, it was implemented by the Allied Forces in 1943 to determine which path their convoys should take in order to minimize interaction with the Japanese.<sup>5</sup>

When applied to turn-based games like Chinese Checkers, the minimax algorithm can maximize the chance

to win a game by minimizing the board positions available to the opponent. The aforementioned algorithm looks ahead into possible future moves by switching perspectives when analyzing the boards. After reaching the final layer of recursive depth in the decision tree, as will be elaborated on later, the algorithm returns the result corresponding to the maximized numerical rating. A key assumption this algorithm relies on involves the opposing player's choice: the algorithm assumes that the opponent will always pick the most optimal choice as dictated by the specific board evaluation routine used in the algorithm. Hence, in some cases, human intuition can lead to humans picking a move that intuitively appears superior to a move that may result in a higher numerical score, the exact formula for which will be discussed in *Implementing Minimax*.

## II. Game Set-upA. Storing the Board

### 1. *make\_board*

The goal of the first main function, *make\_board*, was to store the board positions—including players' pieces and empty board positions—as a two-dimensional array. The board was simplified into a diamond-shaped board for two players starting on opposite sides of the board. This board was then rotated by forty-five degrees counterclockwise, resulting in a nine-by-nine square shape which was then converted into a two-dimensional array with nine rows and nine columns. In the user interface, each position can be accessed and represented by a two-coordinate system using letters A-I for the rows and numbers 1-9 for the columns. For example, the peg at the top left spot, which is notated as location A1, would be accessed in the backend via *board[0][0]*.

As per the aforementioned routine, all ten of player one's pieces were denoted by 1's. When the board was first initialized, player one's pieces were stored in the top left corner of the board, and player two's pieces, which are represented by 2's, were stored in the bottom right opposing region of the board. The remaining 61 empty spots were designated to be 0's in the array's cells. Figure 2 provides a visualization of a starting board position.

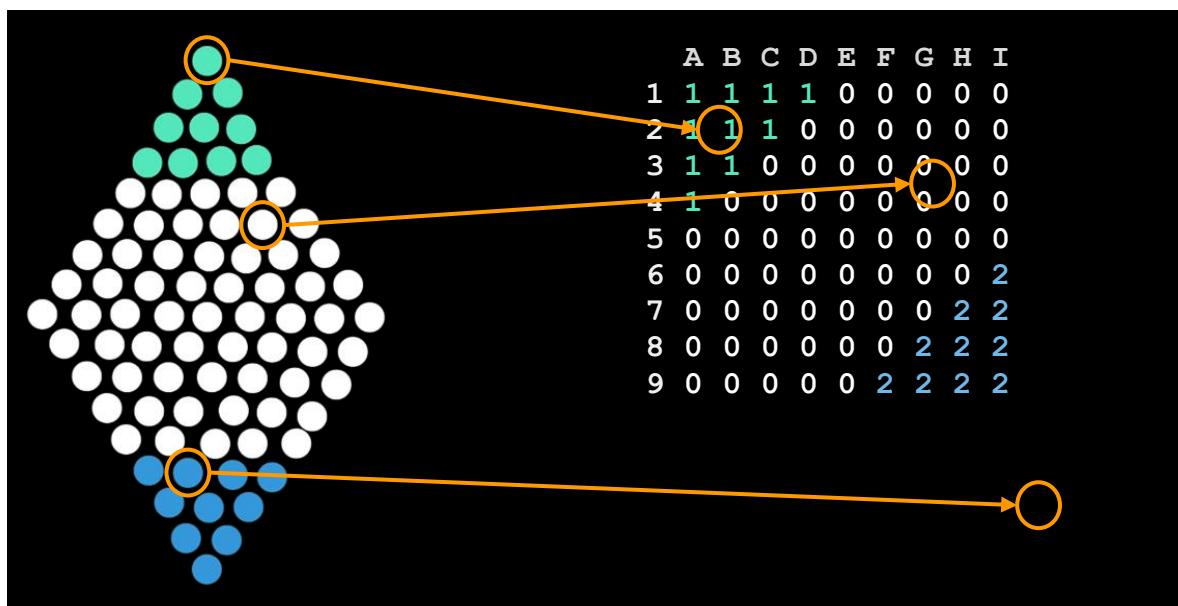


Figure 2: Two-Player Chinese Checkers Board and Equivalent Two-Dimensional Array

## 2. *print\_board*

Although the cells in the arrays contain integers, the board is printed out with numbers at the beginning of each row to denote each line and letters as the top line to denote each column. To avoid confusion with the coordinate labeling system and the actual pegs on the board, the output method converted each of Player 1's pieces from 1's in the array to P's on the display. Then Player 2's pieces were printed as L's on the display, and open spots were represented as the symbol "-". The final user interface is in terms of P's, L's, and -'s with the integer-letter coordinate system, as shown in Figure 3. This labeling system is designed to improve the user interface when interpreting the pegs for a given board position.

	A	B	C	D	E	F	G	H	I
1	P	P	P	P	0	0	0	0	0
2	P	P	P	0	0	0	0	0	0
3	P	P	0	0	0	0	0	0	0
4	P	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	L	
7	0	0	0	0	0	0	L	L	
8	0	0	0	0	0	L	L	L	
9	0	0	0	0	0	L	L	L	L

Figure 3: User Interface of Chinese Checkers Board

## B. Finding Valid Moves

### 1. *validity\_check*

Once the coordinates for a move were acquired, a validity check was run to confirm the move as legal. The *validity\_check* function steps are listed as follows:

1. Confirm that the new position is within the boundaries of the nine-by-nine two-dimensional board by comparing the indices of the coordinates to the indices of the board.
2. Confirms whether the position is vacant by getting the value of the new position cell and verifying that it is a zero, indicating an open cell.

After verifying the legality of the new position, the algorithm must verify the legality of the move itself. As legal moves are either single-spaced moves to adjacent cells or a series of jumps, the checking algorithm compares the differences between the original and new coordinates. For single-spaced changes, it checked whether the change was a vertical or horizontal difference equal to one. Jumping moves were verified by a position difference of two in either a horizontal or vertical direction. Jumping moves also underwent an additional layer of verification to assure a piece was present in the divot being jumped over. In the event that the new move satisfied all of the legal requirements, the *validity\_check* function returned the boolean type true.

## 2. *valid\_moves*

The *valid\_moves* function addressed the need for a list of valid moves from a given board position when creating a decision tree via minimax. For a given player, *valid\_moves* iterated *validity\_check* on all single-spaced and jump moves for each piece. All moves that passed *validity\_check* were appended to a dictionary containing key-value pairs where the keys were coordinates of the pieces and values were lists of valid destination coordinates. Examples of valid moves are depicted in Figure 4.

Possible Single-space Moves  
for Piece at B2

	A	B	C	D	E	F	G	H	I
1	1	1	1	1	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	2
7	0	0	0	0	0	0	0	2	2
8	0	0	0	0	0	0	2	2	2
9	0	0	0	0	0	2	2	2	2

Possible Jump Moves for Piece at A2

	A	B	C	D	E	F	G	H	I
1	1	1	1	1	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	2
7	0	0	0	0	0	0	0	2	2
8	0	0	0	0	0	0	2	2	2
9	0	0	0	0	0	2	2	2	2

Figure 4: Examples of Valid Moves

## C. Random Player

To create a player that makes random valid moves in order to act as a control for the testing phase, a key from the *valid* dictionary was randomly selected using the *random.randrange* method. Then *make\_move* was called to update the board array to reflect the move chosen by the random player. This process will be expanded on further in the *make\_move* section of *Game Play*.

## D. Game Play

### 1. *make\_move*

After a valid move was confirmed, the *make\_move* function updated the array of board positions and the coordinates stored in the player's array of pieces in order to implement the new move. In the array, the original position of the moving piece is replaced with a zero to indicate its updated vacancy, and the new position is replaced with the moving player's number, either a 1 or a 2. The array of pieces is also updated, replacing the old x and y coordinates of the moved peg with the coordinates of its new position. Afterward, when the user interface is displayed the updated piece array is used in order to reflect the current status of the game. The code for the function is shown in Figure 5.

```

def make_move(x1, y1, x2, y2, player, board, pieces):
    try:
        ind = pieces.index([x1, y1])
    except ValueError:
        return False
    board[y1][x1] = 0
    board[y2][x2] = player
    pieces[ind] = [x2, y2]

```

**Figure 5: *make\_move* Function****2. *flip\_board***

The *flip\_board* function rotates the board by 180 degrees, flipping the physical board and updating the player pieces. This allows the same functions to be used on the board without the need of switching the variables. The program flips the board each time it makes a move for the other player and also in *look\_ahead* where the function considers the possible moves of the other player. An example of a board before and after the *flip\_board* function is shown in Figure 6.

	A	B	C	D	E	F	G	H
I								
1	1	1	1	1	1	0	0	0
0								
2	1	1	0	0	0	0	0	0
0								
3	1	0	1	0	0	0	0	0
0								
4	1	1	0	0	0	0	0	0
0								
5	0	0	0	0	0	0	0	0
0								
6	0	0	0	0	0	0	2	0
0								
7	0	0	0	2	2	0	0	0
2								
~	~	~	~	~	~	~	~	~

	A	B	C	D	E	F	G	H
I								
1	2	2	0	0	2	0	2	0
0								
2	2	0	2	0	0	0	0	0
0								
3	2	0	0	0	2	2	0	0
0								
4	0	0	2	0	0	0	0	0
0								
5	0	0	0	0	0	0	0	0
0								
6	0	0	0	0	0	0	0	1
1								
7	0	0	0	0	0	0	1	0
1								
~	~	~	~	~	~	~	~	~

**Figure 6: Board Before and After *flip\_board* is Called****3. *check\_win***

Finally, to know when to end the game, *check\_win* is implemented. As mentioned previously, each player's objective in Chinese Checkers is to completely fill the opposing player's home space with all 10 of their pegs. This, however, is not always possible as it assumes that all of the opposing player's pieces have left their home space. Sometimes, especially when playing against Random, a few opponent pieces may be

left in the goal. Therefore, *check\_win* iterates after each move, checking if all of the 10 holes in a given goal space are occupied and if at least one of the pegs in the goal space belongs to the player, as shown in the examples in Figure 7. If so, the function returns that the opposing player won and ends the game.



Figure 7: Two Possible Winning Configurations for Player 1

## E. Endgame

Near the end of the game, at relatively low levels of search depth, the minimax evaluation function often proved unable to detect a winning move. In these cases, the board evaluation routine prioritizing moves closer to the far corner of the board made the computer hesitant to move pieces backward into lower-scoring positions that would lead to a win. The pieces then became stuck in an L formation at the sides of the board as depicted in Figure 8, alternating between two adjacent holes. This difficulty often caused games to run until the dictated upper bound of 500 moves as any number greater than such heavily indicated that neither player was able to secure a victory. To solve this complication, this research project introduced the endgame algorithm.

A	B	C	D	E	F	G	H	I
1	2	2	2	2	2	0	0	0
2	2	2	2	0	0	0	0	0
3	2	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	1
8	0	0	0	0	0	1	1	1
9	0	0	0	0	0	1	1	1

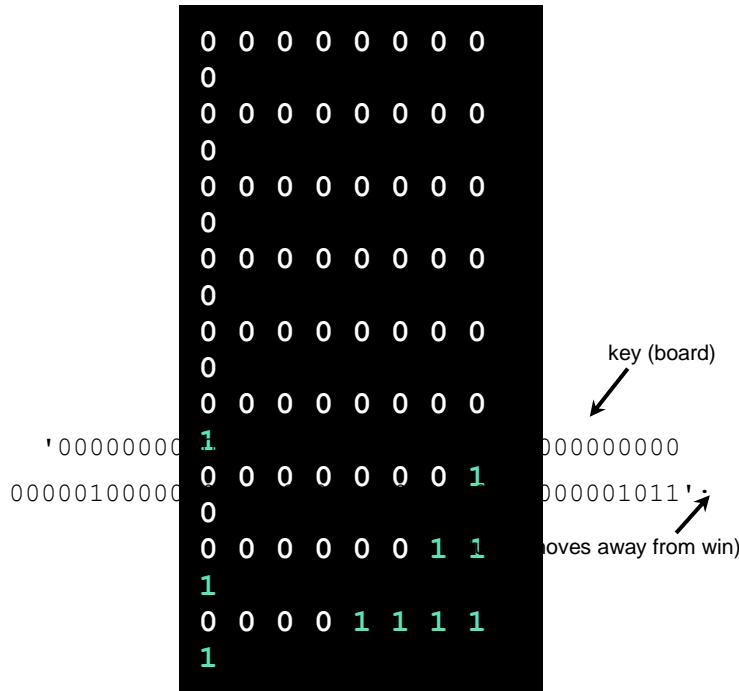
**Figure 8: Complications without Endgame**

### 1. Defining the Endgame

For the purposes of this study, an endgame state was defined as a board for which all of both players' pieces have passed one another. After this stage, each player's moves do not affect the other's moves in any way. Each player thus no longer had to consider their opponent's position, allowing for a simplified endgame calculation and storage approach.

### 2. Storing Solutions

The program corresponding with this research stored all board positions for which the computer is within five moves of winning the game in the appropriate long-term storage file. Each of these board positions were converted to strings and stored as keys in a dictionary. In this structure of key-value pairs the keys are the board positions. The values are the integers corresponding to the number of moves a board position is from the winning position. Figure 9 displays a sample key-value pair from an endgame dictionary alongside the board position it corresponds to. Notice that the string represents the board position and the integer value represents the number of moves said board position is from a completed game.



**Figure 9: Example Element Stored in Array**

Since there must always be a move that brings the position one step closer to completion from each position in the aforementioned data structure, the implementation of the endgame routine became significantly simplified. To elaborate, an algorithm to find the appropriate move given a known endgame position with  $k$  moves left before victory solely needs to search the dictionary for boards resulting from valid moves that contain a value of  $k-1$ .

### 3. Generating the Database: Retrograde Analysis

In order to generate the endgame database, this research relied on the process of retrograde analysis, a method of finding all of the possible positions that could have preceded the current position. A key function implemented in this procedure is *new\_positions*.

#### *new\_positions*

The *new\_positions* function took in two parameters: the string representation of the board and the number of moves away from the winning configuration. Then, the function called *valid\_moves* generated all the valid new positions that are one move away from the inputted position. All of these possible positions were thereafter stored in a list. Subsequently, each element of the list containing a new possible position was checked against the dictionary. If the position was not in the database, the function appended the position to the dictionary with a value one more than the value of the inputted position. Recursively executing this function ultimately created the endgame database containing all the positions within a given number of moves away from completion.

#### 4. Loading in the Dictionary

The endgame database grew exponentially in size as the pieces were farther away from the finished position. Compared to the very end of the game, the middle of a round of Chinese Checkers proved to be much more complex with many more possible moves and strategies to choose from. Table 1 demonstrates this pattern. For each additional number of moves away from a winning configuration, the number of positions stored increased greatly.

**Table 1: Dictionary Increases in Size as Game is Further from End**

Moves Away From Win	Positions Stored	Factor of Increase
0	1	N/A
1	15	15.0
2	163	10.9
3	1,496	9.2
4	11,368	7.6
5	78,000	6.9
6	500,764	6.4
7	2,962,726	5.9

To load in the dictionary, the database generation code was computed externally to generate a text file representing the dictionary. This file was then uploaded to the primary shared file and accessed in the code via the pickle module in Python.

## 5. Implementation: Playing the Game

### ***check\_valid\_endgame***

In order to reconcile the endgame dictionary with the rest of the code, it was necessary to be able to check whether a position is an endgame and could be found in the dictionary. When both of these conditions were met the program would use the endgame routine. In any other case or if the endgame had been disabled, the program would proceed with its typical minimax algorithm. This *check\_valid\_endgame* function thus returned True if the entered position was both an endgame position and in the endgame database. It returned False otherwise, indicating that the game was still in normal play.

### ***make\_endgame\_move***

*make\_endgame\_move* accepted a board position as input that was assumed to be an endgame in the database. The function then returned a move that found a new position in the database that was one move closer to a winning board position by searching through valid moves for the current board position.

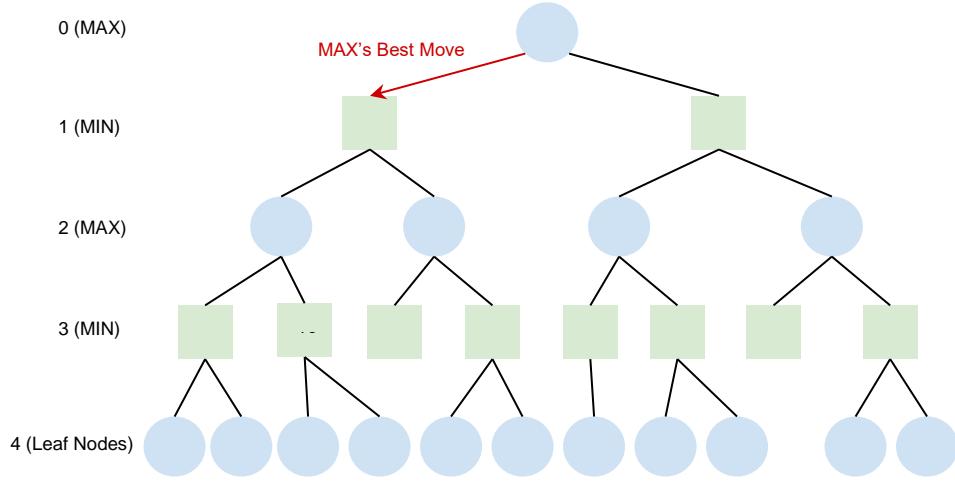
## III. Implementing Minimax

### A. Algorithmic Overview

The minimax algorithm predicts the opposing player's moves given a current board position and assigns each move a score as dictated in *Scoring*. Starting with the current board position, the minimax algorithm loops through all possible moves via the algorithm detailed in *valid\_moves*, subsection B2 of section II. For each of these moves, the minimax algorithm loops through the opponent's responses, considering the game from said opponent's perspective. The algorithm then recursively continues through the computer's response to each of the opponent's moves and so on for a given number of levels of depth.

In the final recursive layer where the levels of depth reach zero, the minimax algorithm measures the board position at the time using the techniques detailed in subsection B. The algorithm assumes that the player chooses the move leading to the highest-scored board position at this lowest level of recursion. Under this assumption, at each upper layer of recursion, the algorithm attempts to select the move that results in the opponent's worst board position. When reaching the top level, the minimax algorithm ultimately chooses the move with the highest score, which also leads to the lowest-scored move for the opponent. During this cycle, however, if any move is found to be an immediate victory for the computer, the function returns a finitely large value, thus incentivizing that move above all others.

A simplified possible decision tree resulting from the minimax algorithm is outlined in Figure 10. As shown therein, each blue node selects the maximum value from the green nodes underneath and each green node selects the node that minimizes the value of its blue daughter nodes. In terms of implementation, as the function is recursive, when running via a green node's perspective the sign of each score is reversed in order to show the most optimal move from the perspective of green nodes. At the bottom of the tree are the different moves that blue nodes could make and their score. In this case, from the perspective of the blue (the minimizer), the AI will choose the most optimal move knowing that the green will choose the most optimal move for them. At the top, it is known that they will pick -8 since it is the most optimal for them (because they are the maximizer), from that the blue player will consider its moves (in this case there are 2 possible moves, 20 and -8). From these two moves, the AI will proceed to consider the next layer, which is the opponent's move, and the most optimal option is 20 (at the bottom of the tree) since the opponent will pick the most optimal move for them which is 20. From that, the AI will proceed to pick 20 in order to secure the most optimal move knowing that the opponent also plays optimally.



**Figure 10: Sample Minimax Decision Tree<sup>6</sup>**

## B. Board Evaluation Methods

*Calc\_board* methods calculate the value of the state of the board given the positions of the pieces. A total of X algorithms were tested. To account for opposing players' positions during the scoring routine, the functions described below followed a call to the *flip\_board* function described in subsection D2 of section II. The board evaluation could thus always consider the perspective of a player originating in the top left corner of the board as displayed on the user interface. Thus, a winning position is defined as pieces all being as close to the I9 position, shown in Figure 7, as possible.

### 1. *calc\_board1*

The *calc\_board1* method is a simple algorithm that only takes into account the pieces on one side of the board. It loops through each piece and sums up the x and y values. It adds the sums, arriving at the final score. The further away each piece is from the starting origin in the top left corner of the board, the higher the score. This encourages each piece to jump as far down and to the right as possible. However, this does not consider how far behind the last piece is from the first piece nor its distance from the diagonal line.

$$\Sigma(x_i + y_i)$$

## 2. *calc\_board2*

Building on the aforementioned algorithm for calculating a board position, the *calc\_board2* method relies on a modified version of the Pythagorean theorem, a commonly used evaluation metric for distance to a given point. In particular, this approach uses the following formula to calculate distance from the end goal:

$$d = (9 - x)^2 + (9 - y)^2$$

By adopting the Pythagorean theorem over a typical summation routine, center squares are deemed closer to the winning position than those on the side, as shown in Figure 11. Since the optimal board position appears to be that which minimizes the distance to the winning position and the typical Chinese Checkers strategy incentivizes moving pieces to the center, this routine to calculate the board can be expected to increase the overall minimax performance. The distance value calculated above is then subtracted from an arbitrarily large constant, supposed as 1,000 in the case of this approach, for a higher final score to indicate a better board position.



**Figure 11: Distances as Evaluated by *calc\_board2* Algorithm for Player Starting in Top Left Corner**

## 3. *calc\_board3*

The *calc\_board3* function uses a combination of distance and standard deviation to calculate a score. It utilizes the same distance function as in *calc\_board2*, but also subtracts off a weighted standard deviation. In Chinese Checkers, it is generally good for pieces to stay relatively close to each other so that longer chains of jumps can be made. Therefore, having the standard deviation of the x and y values be small is a benefit to the score.

$$1000 - \Sigma((9 - x)^2 + (9 - y)^2) - \sigma_{p1}(weight_{p1}) - \sigma_{p2}(weight_{p2})$$

#### 4. calc\_board4

*Calc\_board4* uses the distance from the end goal triangle and from the central diagonal line for its score calculation. The central diagonal line is indicated below in Figure 12. *Calc\_board4* is marginally similar to *calc\_board3* but uses a different method of maintaining the pieces in a line. To start, *calc\_board4* finds the difference between the x and y values. Then, it subtracts that weighted difference from the total score. This weight was fine-tuned for the best results. In this way, pieces are encouraged to aim for a smaller difference in the x and y values, thereby keeping them closer to the central diagonal line. In Chinese Checkers, this prevents pieces from straying too far from each other or running along the edge into the wrong corners, which is very inefficient.

$$1000 - \sum((9-x)^2 + (9-y)^2) - |x - y|$$

	A	B	C	D	E	F	G	H	I
1	P	P	P	P	0	0	0	0	0
2	P	P	P	0	0	0	0	0	0
3	P	P	0	0	0	0	0	0	0
4	P	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	L	
7	0	0	0	0	0	0	O	L	L
8	0	0	0	0	0	L	L	L	L
9	0	0	0	0	0	L	L	L	L

Figure 12: Board Showing Central Diagonal Line

#### C. Pruning

A significant constraint associated with the traditional minimax algorithm is the runtime of  $O(10^n)$ , for the levels of search being  $n$ . As shown in Figure 13, even marginal increases in the levels of depth lead to almost unreasonably large runtimes given the computing resources available for this research. Hence, techniques to minimize the runtime for each level of depth would allow the algorithm to create a minimax tree that is deeper than those otherwise possible with the same resources and runtime.

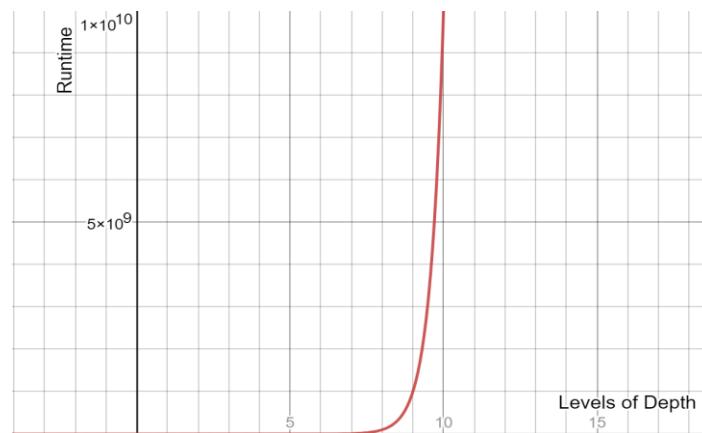


Figure 13: Runtime vs. Levels of Depth  
Graph made using Desmos

To elaborate, the pruning approach used for this project adds an additional parameter to the function designed to look at future turns, starting with that of the opposing player. This parameter represents the ideal move at any given time in the loop. If the recursive state then comes across an option for the opponent that would yield a worse position for the computer than the best one previously found, this addition “prunes” that node of the tree and eliminates recursion therein. An example of a pruned branch is shown in Figure 14. Through this approach, the minimax algorithm increases efficiency and prevents the needless exploration of branches that have no impact on the win rate of the program, as will be proven via testing.

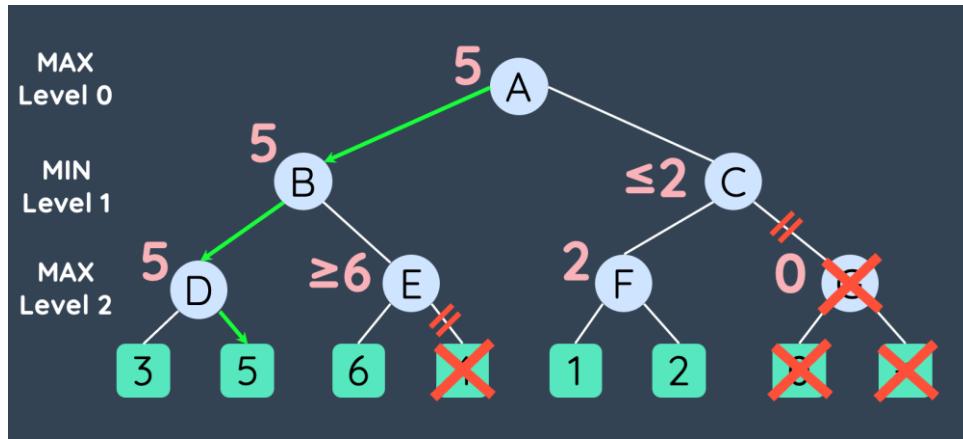


Figure 14: Example of Alpha-Beta Pruning

## IV. Testing

To compare the efficacy of the three *calc\_board* functions and different levels of *look\_ahead* with the Minimax function, this research included different tests and compared the percentage of times each method won and the amount of turns it took for a player to win the game. This study also recorded the percentage of trials that ended as draws, where players 1 and 2 won in the same number of moves, and the percentage of trials that were inconclusive as they used over 500 moves. These inconclusive trials are expanded on in section A2.

### A. Controls

#### 1. Minimax Player vs. Random Player

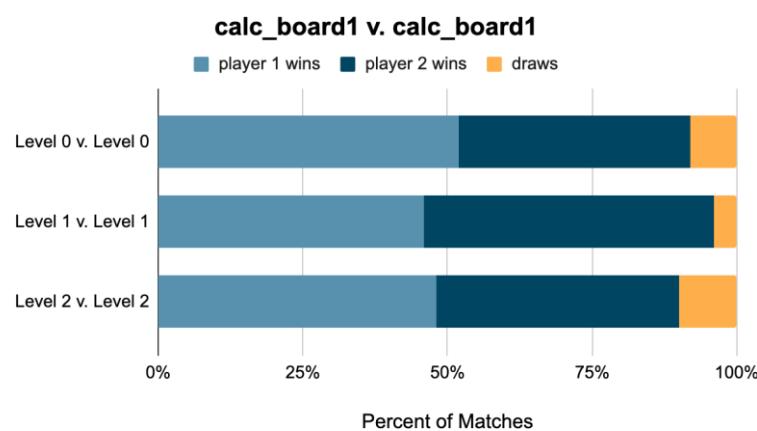
The first trials of this experiment involve the various level-zero *calc\_board*s run against a random player. These tests were initially run to illustrate the raw playing capabilities of the most primitive algorithms. Each of these *calc\_board* v. random players executes 100 matches of Chinese Checkers. The consistently high winning percentages demonstrate the program’s superiority over the random player. As Table 2 indicates, all of the *calc\_board* methods are able to understand the game and strategize moves at a basic level.

**Table 2: Percentages of Matches Won by Random v. calc\_board Routines**

	calc_board1	calc_board2	calc_board3	calc_board4
% of trials won by calc_board1	100%	100%	100%	100%
% of trials won by random player	0%	0%	0%	0%
% of trials that ended as draws	0%	0%	0%	0%

## 2. Identical Intelligence Tests

In order to establish a baseline for the next experiments, this research included 50 round trials of Chinese Checkers for *calc\_board1*, *calc\_board2*, and *calc\_board3* against themselves. When doing so, one can expect that the numbers of wins and losses for each player would be rather equal, considering that the two opposing players should be equivalent in playing ability. The control group results for *calc\_board1*, exhibited in Figure 15, support the notion that the *calc\_board1* strategizing algorithm remains relatively similar in level when playing against itself. However, the results for *calc\_board2* are not as clear. Although the even levels (level zero and level two) give some support for the board's consistency, the odd level (level one) shows a large portion of results being stalemates between the players. This outcome likely resulted from the fact that odd-level boards play defensively, as will be discussed in more depth in the *Discussion*. When the players play entirely defensively instead of offensively, they tend to avoid each other, spreading out to the sides of the board in "L" formations. The players will avoid each other at all costs to prevent the other player from jumping over their pieces. This situation causes both players to stray to the sides of the board, creating an infinite stalemate loop as can be seen in Figure 16.



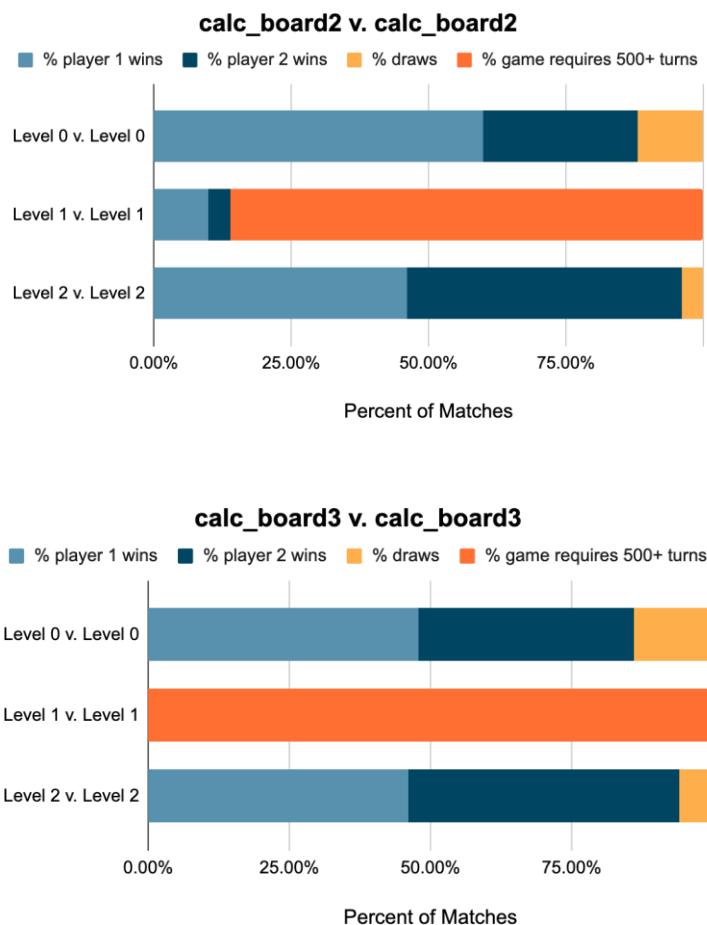


Figure 15: Identical Intelligence Tests for calc\_boards 1-3 and Levels 0-2

	A	B	C	D	E	F	G	H	I
1	2	2	2	2	2	0	0	0	0
2	2	2	2	0	0	0	0	0	0
3	2	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	1	1	1
9	0	0	0	0	0	1	1	1	1

Figure 16: Endless Game Example With “L” Formation

## B. Comparing Board Evaluation Methods

### 1. Overall Comparison

To compare the `calc_board` metrics, the different metrics were played against each other at level zero. Each comparison was a result of 50 trials of Player 1 v. Player 2 at Level 0.

**Table 4: Comparison of calc\_board1, calc\_board2, calc\_board3, and calc\_board4 at Level 0**

		Player 1				
		calc_board1	calc_board2	calc_board3	calc_board4	
Player 2		calc_board1	X	X	X	X
calc_board2		0% / <b>100%</b>	X	X	X	
calc_board3		0% / <b>100%</b>	38% / <b>56%</b> / 6%	X	X	
calc_board4		2.08% / <b>97.92%</b>	32% / <b>60%</b> / 8%	<b>48%</b> / <b>46%</b> / 6%	X	

In the above table, the percentage of matches that Player 1 won is listed first, then the percentage of matches that Player 2 won, and lastly the percentage of draws if applicable. These percentages are separated by /'s.

The results showed that `calc_board2` and `calc_board3` overwhelmingly outperformed `calc_board1` in 100% of the trials. As for `calc_board1` v. `calc_board4`, `calc_board1` did win one match against `calc_board4`. However, this data point is an outlier among the 50 trials and could have just occurred due to chance from the small sample size. `calc_board4` also significantly outperformed `calc_board2` but with a smaller margin than `calc_board1` v. `calc_board4`. Although `calc_board3` also outperformed `calc_board2` in most trials, the winning lead was relatively minimal. This outcome was likely due to the fact that `calc_board2` and `calc_board3` both relied on the distance value described in section B3 of *Implementing Minimax*. This indicates that the additional standard deviation value used in `calc_board3` may have not led to as large of an advantage as the distance value did. Lastly, the results from the `calc_board3` v. `calc_board4` trials were inconclusive, likely due to the metrics' similarities discussed in section B4 of *Implementing Minimax*.

### 2. Comparing calc\_board1 and calc\_board2 across 5 levels

Since the results `calc_board1` v. `calc_board2` were quite unique in that `calc_board2` won against `calc_board1` in every trial, a more comprehensive analysis was conducted looking at results for more levels of `look_ahead`. When `calc_board1` played against `calc_board2`, `calc_board1` won the majority of the time for odd levels of `look_ahead` in the minimax and lost most of the time for even levels. It is possible that `calc_board2` played better offensively as when both metrics were playing offensively for the even levels, `calc_board2` had a higher winning percentage than `calc_board1` did. However, `calc_board1` appeared to play better defensively, which is shown in the odd levels. When the algorithm looks ahead an odd number of levels, the last move analyzed is that of the opponent, so the opponent's moves are weighted more heavily. Since the computer assumed a defensive stance, it often avoided moving its pieces in the middle to prevent building a bridge that the opposing player could use to jump over.

**Table 5: calc\_board1 v. calc\_board2 for Levels 1-5**

	Level 1	Level 2	Level 3	Level 4	Level 5
% of trials won by calc_board1	<b>98%</b>	8%	<b>100%</b>	0%	<b>98%</b>
% of trials won by calc_board2	0%	<b>90%</b>	0%	<b>100%</b>	2%
% of trials that ended as draws	2%	2%	0%	0%	0%

### 3. Comparing calc\_board2 and calc\_board3 for Levels 1-5

When *calc\_board2* played against *calc\_board3*, *calc\_board3* won most of the time in the even number levels. This is likely due to the fact that it considered the standard deviation while *calc\_board2* did not. However, the game was usually inconclusive for odd levels. When the game ran for over 500 turns, the program marked it down as such and moved on to the next trial. This happened when both players' pieces got stuck in an "L" formation on the corners of the boards. They were unable to activate the endgame algorithm since the pieces were too far away from the winning position. Additionally, they were also too short-sighted to win the game. This will be expanded on in the subsection of *Discussion* titled "Endless Games." The inconclusive games occurred for the odd number levels since the players both played too defensively, moving pieces to the edges of the board to avoid helping the opponent in the middle. Therefore, *calc\_board3* beat *calc\_board2* for even levels of look ahead, and games were inconclusive for odd levels of look ahead.

**Table 6: calc\_board2 v. calc\_board3 for Levels 1-5**

	Level 1	Level 2	Level 3	Level 4	Level 5
% of trials won by calc_board1	20%	46%	14%	30%	16%
% of trials won by calc_board2	0%	<b>52%</b>	0%	<b>68%</b>	0%
% of trials that ended as draws	0%	2%	0%	2%	0%
% of trials that took 500+ turns	<b>80%</b>	0%	<b>86%</b>	0%	<b>84%</b>

### C. Endgame

For the next area of data collection, the program ran a number of trials to isolate the endgame's advantage in strategizing a series of ending moves to win the game. The addition of the endgame in the program significantly increases the computer's ability to win at Chinese Checkers. As seen in Figure 17, when *calc\_board1* with endgame is run against *calc\_board1* without the endgame, the player with the endgame wins about three times more often than the player without the endgame. These results show how implementing endgame drastically improves the ability of the computer to win in Chinese Checkers. Another observation is that as the levels increase, there is a slight upward trend in the winning percentage of the player with the endgame. However, this pattern is not significant enough to indicate that there is a correlation between higher levels and a larger gap between the more efficient endgame player and the player without endgame.

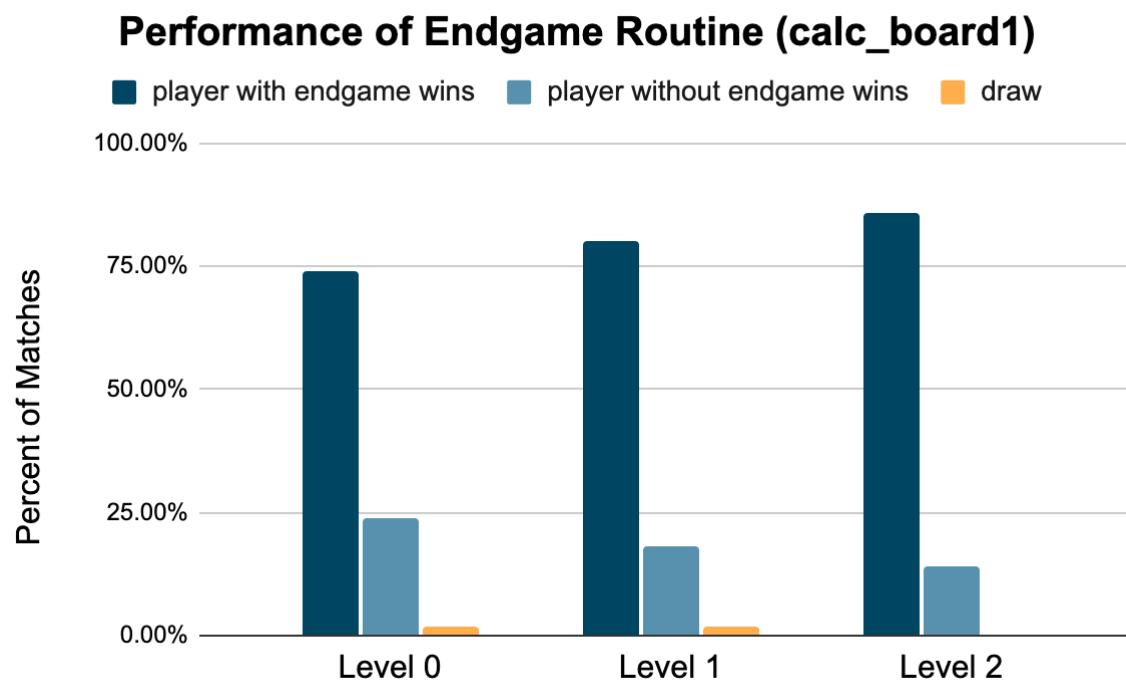
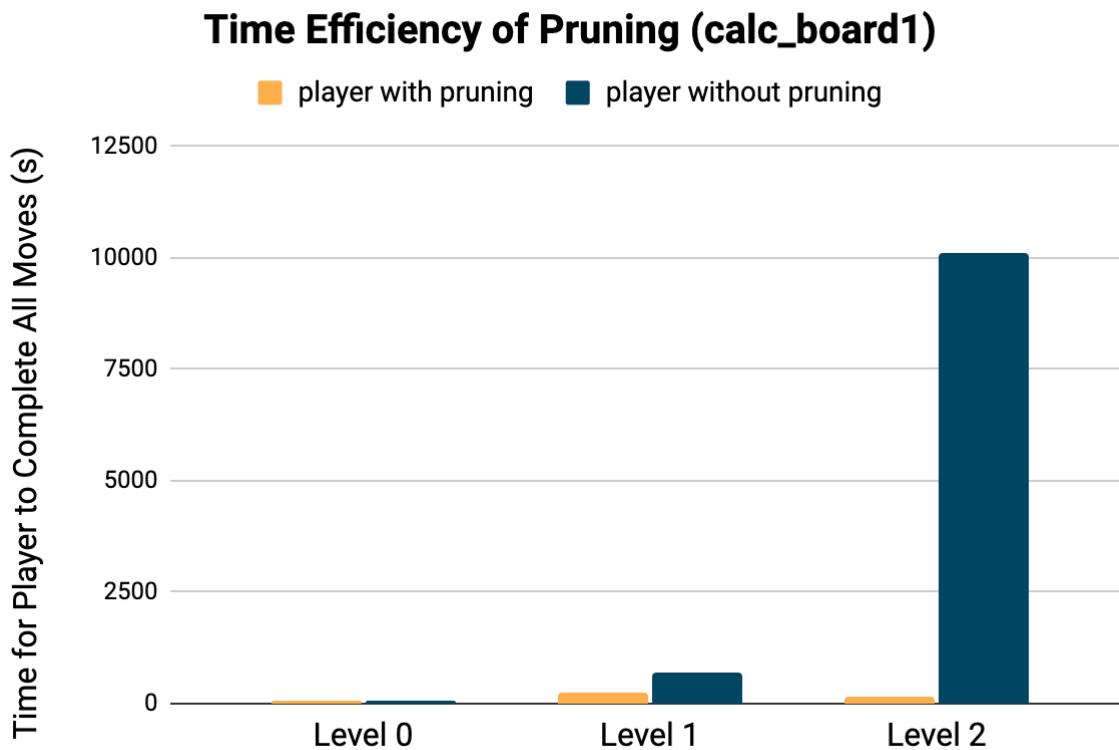


Figure 17: Performance of Endgame v. No Endgame (*calc\_board1*)

### D. Alpha-Beta Pruning

Our next component of testing evaluates the time efficiency of Pruning for *calc\_board1*. Although pruning has very little effect on run time for level zero, as the levels increase, the time efficiency of pruning becomes exceedingly apparent. This considerable gap is illustrated in Figure 18. Based on these results and the process behind pruning, one can infer that pruning becomes even more valuable in allowing time-efficient rounds for games involving a large number of layers.



**Figure 18: Performance of Endgame v. No Endgame (*calc\_board1*)**

## V. Conclusion

### A. Discussion

#### 1. Short Games with Random Player

In games involving the random player, the run time consistently proved to be faster with an unusually low number of moves leading to a win. Further exploring these cases, the aforementioned observation can likely be attributed to the random player leaving some pieces in their starting positions. While this is intrinsic to tests conducted against a random player, further tests against more intelligent algorithms would likely prove less definitive and distinctive results. Hence, while this “vs. random” test remains a viable baseline test, it is likely an incomplete representation of each algorithm’s potential.

#### 2. Endless Games

Without the endgame algorithm, the *calc\_board* levels 0, 1, 2, and 3 often were unable to see far enough into the future to finish the game. This is because pieces on the edge often need to jump out to a position with a lower-rated board position or take a route that seems harmful in the short run despite leading to a winning move after a series of turns. Without looking a sufficient number of levels deep, however, the algorithm is unable to detect the aforementioned temporary sacrifice and gets stuck in a loop of jumping back and forth between adjacent spaces.

While the endgame algorithm typically addressed this issue, some select endless games remained endless. In particular, the endgame routine currently adopted for this research was only active when the game was within five moves of a winning position. Thus, in scenarios where pieces were in an “L” position such that a winning move was six or more moves from a winning position, the endgame remained unactivated.

In addition, when playing against the random player, the pieces of one player often had not completely passed those of the opposing player. Thus, in such cases, the conditions needed for the endgame algorithm were not met. During these cases when pieces from one player stayed in their starting positions, the intelligent player benefitted at level 0, where the intelligent player was only considering the next move for its own pieces. However, levels 1, 2, and 3 each assumed that the random player would make smart, calculated moves. Therefore, the algorithm was more likely to move pieces out to the edges of the board to avoid providing opportunities that helped the random player. This often resulted in further never-ending loops.

### 3. Even vs. Odd Levels

When the levels of lookahead were even, the player tended to play more offensively as the final layer of the minimax algorithm was looking at the player’s own moves rather than the opponent’s move. On the contrary, when the number of layers was even, the algorithm tended to play more defensively. To elaborate, defensive play indicates that the pieces tend to distance themselves from the opponents’ pieces to avoid helping them. The pieces thus tended to stray towards the edges of the board, sometimes getting stuck in an L-formation not present in the endgame database, leading to an endless game as described in the *Endless Games* section above.

### 4. Comparing Calc\_boards

*Calc\_board1* and *calc\_board2* both solely consider distance. However, at level zero, the higher the version of *calc\_board*, the better it performed. The only exception to this trend was evident when comparing *calc\_board3* and *calc\_board4* as they performed with similar efficiencies. Such an anomaly in the trend was expected, however, as *calc\_board3* and *calc\_board4* both approach prioritization of centered board positions via different formulae.

## B. Limitations

### 1. Computational Power

The algorithm outlined in this paper was largely limited by the computational power available via Repl.it and local devices used via Visual Studio Code. Additionally, using more advanced processing algorithms—like using more levels of search or more complex scoring methods—required far more time to run, constraining the research to relatively basic algorithms. As reflected in Figure 13 discussed earlier, each additional layer of search depth in the minimax algorithm increased the runtime exponentially; thus, even when considering pruning, the number of layers of search depth in the minimax decision tree is significantly constrained.

## 2. Board Scoring Algorithm

Due to time constraints associated with this research project, fine-tuning for the board evaluation routines was imperfect. In particular, when fine-tuning weights of different factors considered for board evaluation, algorithms with different weights were played against one another; however, this process was conducted manually with weights chosen by the human operator. Hence, the truly optimal weight may have been between two tested weights, which could have only been caught by an automated weights calculator.

As for the board scoring routines themselves, since they were constructed based on human observations rather than a more concrete automated technique, there may have been other scoring routines performing better or with a faster runtime than those adopted for this study.

## 3. Time Evaluation

Since there may have been minuscule differences in the available operating power during the testing process for each pair, the time differences evaluated for pruning may have been slightly affected by ongoing background processes. However, this difference is likely to have been negligible as each player's time was calculated at approximately the same time on the same device, allowing for the conclusions listed earlier.

## 4. Small Sample Sizes

Although the sample sizes for data considered in this research were not incredibly small and were suitable given the tight time constraints of this research, some of the results seemed too close to draw definitive conclusions with solely 50 samples. If this research had instead adopted a larger sample size of  $n = 500$  or  $n = 1,000$ , more definitive conclusions may have been plausible.

## 5. Time to Load in the endgame table

When adopting the endgame table, since only board positions within five moves from a winning position were considered, some endgames continued to cause endless games. Although part of these issues may have been accounted for if the research relied on the endgame dictionary generated up to seven moves away, such a file took longer than ideal in preliminary trials that were not reported in this study.

# C. Future Work

## 1. Human Player vs. AI Player

Trials presented in this paper primarily focused on the computer vs. computer Chinese Checkers games. Since the purpose of using the minimax algorithm is to apply its potential to real-world situations, a next step worth researching could be human vs. computer rounds. For the purposes of this project, as the sample size would have been too small to draw conclusions, such tests were not conducted; however, future research with additional time and resources could also report statistics representing the computer's win rate against human players with varying levels of skill.

## 2. Machine Learning

While the minimax approach can predict a computer's moves with a high level of accuracy, human players are often more complicated. For instance, the minimax algorithm does not share humans' "bird's eye" view of the board or retain memories of previous games, thus causing it to make moves that seem unintuitive to

humans. To account for such, a machine learning approach could modify the value associated with each board position based on previous games wherein the computer lost despite making the best move dictated by minimax.

Further improving on the aforementioned machine learning approach, the computer could follow the path laid by well-known models like AlphaGo<sup>7</sup> and adopt a deep learning approach. This approach includes creating neural networks in the program that allow it to learn the game via experimentation in a human-like way. By training such a computer model against itself, it would likely identify strategies that the original creators may have been unaware of and avoid disadvantageous moves in the long run. This addition would likely improve the computer's performance dramatically and is thus a practical avenue that future research should investigate.

### **3. Deeper Layers of *look\_ahead***

Future research that faces fewer limitations in the realm of processing power could search deeper layers in the recursive minimax tree. These additional layers allow the computer to take additional moves into consideration when choosing an immediate upcoming move. With each new layer of the decision tree, the program can be expected to approach and possibly exceed human intelligence for the art of Chinese Checkers.

### **4. Combining Endgame with *look\_ahead***

The endgame dictionary currently adopted in this search encompassed only a portion of all possible endgame positions. In order to extend the reach of the program beyond the depth of the dictionary, it is possible to combine endgame with *look\_ahead*. This adaptation would allow the program to find the optimal solution to positions a few moves out of range of the dictionary. For instance, during each recursive iteration of the *look\_ahead* recursive function, such a modified approach could check for an endgame position. With such an approach, even if a winning position was just outside the reach of the minimax tree, the algorithm would likely accurately select the quickest path to a winning position.

### **5. Combining Endgame with minimax to find forced wins**

If the program had the ability to store a much larger endgame database and could run deeper layers of lookahead, it would be possible to look ahead from middle game positions to find any moves that force a winning endgame position.

## **VI. Acknowledgments**

We are deeply thankful to everyone who has made this research possible by not only providing subject matter expertise but also motivating us to persevere through bugs and unexpected results alike. In particular, we would like to thank our faculty advisor, Dr. Neil Simonetti, for guiding us through the process of implementing Minimax in our program. We are also truly grateful for our teaching assistant, Liam Hower, who has helped us tremendously in clearing up bugs and supporting us through this journey. We would also like to thank Dr. Barry Luokkala, the Pennsylvania Governor's School for the Sciences Alumni Association, and Carnegie Mellon University for providing us with the opportunity to conduct this research.

## VII. References

1. Willingham, E. (2023, March 13). *AI's Victories in Go Inspire Better Human Game Playing*. Scientific American. Retrieved July 31, 2023, from <https://www.scientificamerican.com/article/ais-victories-in-go-inspire-better-human-game-playing/>
2. Chen, J. Su-I, L. Vekhter, D. (1999). *Strategies of Play*. Stanford University. Retrieved July 31, 2023, from <https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/Minimax.html>
3. Sweetooth Design Company (2020). *CHINESE CHECKERS GERMANY 19TH CENTURY*. Sweetooth Design LLC. Retrieved July 31, 2023, from <https://www.sweetoothdesign.com/games-chinese-checker>
4. Colman M., A. (1982). *4 - Two-Person, Zero-Sum Games*. Science Direct. Retrieved July 31, 2023, from <https://www.sciencedirect.com/science/article/pii/B9780080260709500096>
5. Gamesver Team (2023). *History of Chinese Checkers: 20 Facts (Origins, Controversies,...)*. Gamesver. Retrieved August 1, 2023, from <https://www.gamesver.com/history-of-chinese-checkers-facts-origins-controversies/>
6. Hock-Chan, C (2012, May). *Java Graphics Tutorial Case Study on Tic-Tac-Toe Part 2: With AI*. Nanyang Technological University. Retrieved August 2, 2023, from [https://www3.ntu.edu.sg/home/ehchua/programming/java/javagame\\_tictactoe\\_ai.html](https://www3.ntu.edu.sg/home/ehchua/programming/java/javagame_tictactoe_ai.html)
7. *AlphaGo*. Directed by Greg Kohs, Netflix, 21 Apr. 2017

## Appendix A: GitHub

The code written for this project can be found at <https://github.com/ParthParikh04/Chinese-Checkers-PGSS>

## Appendix B: User Interface

### A. Implementing User Moves

The goal of the program was to create a playable game of Chinese Checkers between a human user and a computer, so it was necessary to adapt the game to make an interactive user interface.

#### 1. *get\_og\_position*

The *get\_og\_position* function prompts the user to enter the coordinates of the peg they would like to move. Since the row coordinates are in terms of integers 1-9 and the column coordinates are in terms of letters A-I, the user is expected to enter a string of length two where the first character is a character within the A-I bounds and the second character is an integer between 1-9. If the length of the string entered is not equal to two, the user is asked to enter a new set of coordinates. After an appropriate-length string is received, the function then converts the letter to its equivalent integer index in the board array using the *convert\_letter\_to\_num* function. The resulting coordinate returned contains the indices of the cell corresponding to the piece to be moved.

#### 2. *get\_new\_position*

Similar to the function getting the original peg position, the *get\_new\_position* function takes in a set of letter-integer coordinates and verifies the length is valid. Subsequently, it converts the two-character string to a set of two integers representing the array cell where the user would like to move their piece.

### B. Checking User Moves

#### 1. *user\_change\_piece*

When involving a human user in the game, the algorithm must recognize that human error is inevitable. In the event that the user enters an invalid move, the program asks the user whether they would like to try another move with the same piece or switch to a new piece to move. This user prompt is shown in Figure 19.

Sorry, that is not a valid move. Please try again.

If you would like to move somewhere else, enter 1. If you would like to select a new piece to move, enter 0.

**Figure 19: *user\_change\_piece* Prompt to User**

## ***2. continue\_move***

The continue\_move function gives the user the freedom to control the number of jumps they make when a chain of jumps is possible during their turn. By processing the user's response, the program either prompts the user to enter the coordinates of their next jump or it terminates their turn. The continue\_move function is called after each jump. The user prompt for this check is displayed in Figure 20.

Would you like to continue jumping? Enter 2 for YES and 3 for NO:

**Figure 20: continue move Prompt to User**



# **Connecting the Dots: Using Turn-Based Game Strategy to Solve and Boxes**

Angela Abraham, Jonathan Barsotti, Noah Beckish, Karina Ker-Wei Chan-van der Helm,  
Yeana Kim, Kamyra Rajesh

## **Abstract**

The Dots and Boxes team project focused on three aspects of the game: creating a python-based program to play the game, finding the most efficient strategy to win the game, and implementing the strategy to allow a computer to beat an average player every time. The main focus of the project soon shifted to optimizing computational time versus “perfection” of the program. This report discusses our approach to building a computer program that finds the best set of moves that can be used to win a game of Dots and Boxes against an average player. We implemented both the minimax algorithm and alpha-beta pruning in order to find optimal moves and try to reduce computational time. We also collected and analyzed data from our computer playing against humans and other computers to understand the importance of certain strategies in the game. We concluded that our code could definitively beat the average human the majority of the time, showing the value of the minimax algorithm and a strategy called double-crossing that was found during our research. We were also able to conclude that going first in the game gives the first player a significant advantage.

## **I. Introduction**

### **A. Dots and Boxes Overview**

Dots and Boxes is a 2 player game that is played on a grid of dots of any size. The players take turns connecting adjacent dots using either a horizontal or vertical line. Diagonal lines and lines that span more than 2 dots are not permitted. A box is defined by 2 horizontal and 2 vertical lines that connect adjacent dots in a square shape, with each side length being the distance from one dot to an adjacent dot. The objective of Dots and Boxes is to end the game with more claimed boxes than the opposing player. The player that draws the fourth/last line of a box claims it regardless of who played the previous box lines. If a player makes a move that completes a box, they must make another move. The game ends when there are no legal moves left.

### **B. Strategy**

The first few moves are not important to the game's end since a handful of scattered lines doesn't allow for many game-determining moves where a player can create numerous boxes. In the following sections, we will talk about how to play (and win) Dots and Boxes. We will consider the “greedy” algorithm where the opponent tries to take as many boxes as possible, and also the case where surrendering some boxes is necessary to eventually claim more boxes than an opponent.

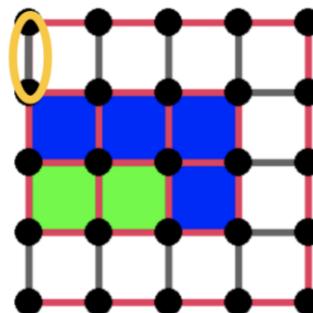
#### **1. Basic Strategy**

The most basic strategies of Dots and Boxes include looking for lines that complete boxes and avoiding making lines that complete the third line in a box. Making the fourth line in a box and completing it is

advantageous because it adds to the player's score. But making the third line in a box should be avoided because it allows the opponent to make the fourth line of the box and then take the box for themselves.

## 2. Cascades

A cascade is a chain of 3 or more boxes that can be taken in a single turn. Players can either initiate or claim a cascade. A player initiates a cascade by drawing the third line of a box, which then allows an opponent to draw in the fourth line of a box and claim it. In a cascade, the fourth line, granting the player an additional move, acts as the third line of another box which can then be claimed with one line, granting yet another additional move, and so on to the end of the cascade.



**Figure 1: Cascade example**

In Figure 1, whoever plays the move circled in yellow would initiate the cascade by drawing in the third line of the box. Then, the next player would be able to claim the cascade by completing every single box left on the game board.

A player should avoid initiating cascades and should instead aim to force their opponent to initiate cascades; when an opponent draws a line that initiates a cascade, the other player may claim the cascade and win multiple boxes in one turn. In cases where there are multiple cascades on one game board, a player may look to concede a few boxes or the shorter cascade in order to eventually secure the longer cascade and win the game. A rule of thumb is that it is in the player's best interest to have no non-cascade initiating moves at the end of their turn. This forces the opponent to initiate the first cascade.

## 3. Double-crosses

To an inexperienced player, making moves that don't allow the opponent to claim a box and likewise, claiming the most boxes possible in a single turn seems to be the best strategy. However, a deeper analysis of the game reveals a frequently found situation: a two-box chain that is key to a player's control of the game. A player can utilize this situation by giving up those two boxes in order to force their opponent to initiate a cascade. Figure 2 shows an example in which if a player does not double-cross (claims the two boxes), they are forced to make an additional move. Double-crossing is useful when a player wants to change who initiates the first cascade. The player would want to double-cross when that additional move initiates a cascade, as the player can then ensure that they win all of the boxes in the cascade. When claiming this cascade, the player can then repeat the strategy, once again sacrificing the last two boxes to ensure they get another full cascade. This cycle can go on as long as those two sacrificed boxes do not give the opponent the winning number.

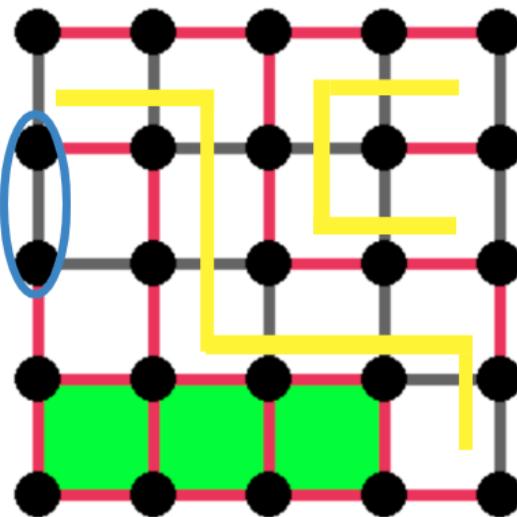


Figure 2: Double-crossing Example

## II. Game Theory and Turn-Based Games

Game theory applies mathematical techniques to analyze situations where players make interdependent decisions; it is the study of competition between opponents. These situations cause the need for each player to take complex consideration of possible moves and their implications for both themselves and their opponents. Each player needs to consider the intentions of their opponents, which could be the same as the player, opposing the player, or some combination of both.<sup>2</sup>

### A. Implications of Game Theory

While games are ultimately sources of entertainment, the same ideas, strategies, and processes that are used can be applied in various situations where parties make choices that are dependent on each other. Game theory optimizes decision making and can be used for any competitive situation including pricing competitions, management negotiations, auction bidding and military conflicts. One important real-life example of this is economics. In certain ways, decisions made in economics can be considered as game theory, as people make decisions based on the choices of other people, trying to optimize their own personal gain. Optimal decisions in economics must be made by analyzing and anticipating moves from both sides just like in a game.<sup>2</sup>

### B. What is a Turn-Based Game?

A turn-based game refers to a game where players take turns to play moves, allowing for pauses in a game, where players can deliberate the best moves to make. Some examples of turn-based games include Chess, Connect 4, Tic Tac Toe, and Checkers.

### C. Board Scoring

Every time a move is made in a Turn-Based Game, the game board is altered. Each game board can be evaluated in terms of its favorability, or how close it is to a winning configuration, for a given player and

assigned a numerical score. A single game may have multiple intuitive ways to score the board, as there are numerous strategies that can be used to play a game, making various configurations advantageous for differing reasons. The more highly scored board configurations are closest to winning board configurations. A program that is designed to play optimally would need to take these scores into consideration when determining ideal board configurations that it would like to aim to achieve. Given the ability to calculate the scores for the possible boards, the next step would be finding the moves that the program would need to make in order to get to these configurations. This requires an advanced algorithm.

## D. Minimax Algorithm

The Minimax algorithm implemented in our Dots and Boxes code prompts the computer to look for the most optimal move to play in any given position. This is done using a board scoring mechanism to compare future positions. The computer looks ahead and compares the outcomes of the various moves. In order to do this, the computer builds a tree structure, with multiple branches for each level of thinking. This is called a decision tree; an example of which is shown in Figure 3. Once the tree is built, each end state of the game is assigned a value based on the game's board scoring algorithm. Then, the algorithm evaluates these scores by comparing them to each other and choosing one score to send back up the tree to the prior level. The score that is chosen to be sent up the tree varies depending on whose turn it is. One player attempts to maximize their score (maximizer), while the other tries to minimize the score of the first player (minimizer). The decision tree is in the point of view of the maximizer, so on its turn it will choose the move with the highest score (best possible move for itself) while the minimizer is assumed to play the move with the lowest score (best move for the minimizer and worst move for the maximizer). The algorithm does this because it makes the assumption that the minimizer is playing optimally. This tree essentially gives the computer the worst-case scenario by forcing it to consider the minimizer's optimal moves and then play the best move it can from there. If the minimizer fails to play optimally, then the computer automatically gains an advantage. If the minimizer does play optimally, then this algorithm allows the computer the chance to still beat its opponent. Positive score values in the tree indicate a winning move for the maximizer, while negative scores indicate a winning move for the minimizer.

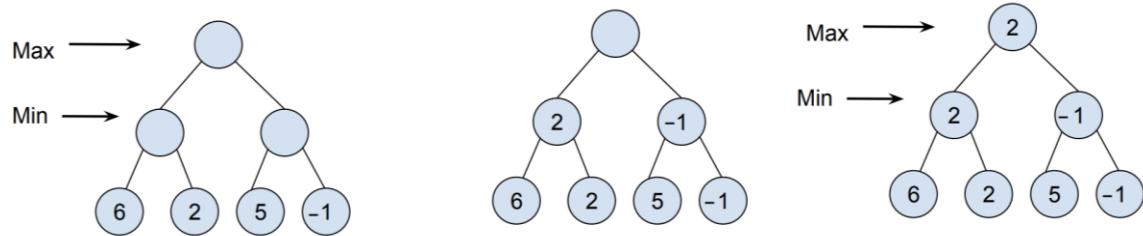


Figure 3: Minimax Decision Tree

While the strategies for playing Dots and Boxes can be taught to a human player relatively quickly, programming these strategies into the computer can be quite difficult. But, in theory, implementing the minimax algorithm allows the computer to play the best move without being made aware of all of the strategies of the game. It is simply able to look ahead at the boards that would be created from each possible move, decide what the best board is based on the board scoring function, and work its way back up the tree to determine the best move to make.

## E. Alpha-Beta Pruning

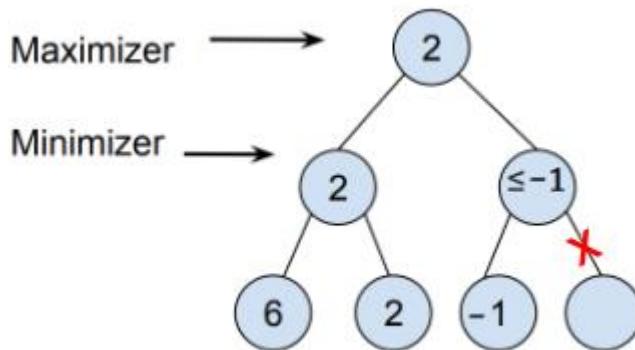
A disadvantage of the minimax decision trees is the exponential rate at which the number of nodes grows. For example, in Dots and Boxes, the number of branches on the tree structure depends on the number of available moves left. At first, there are 40 moves on the 16-box (25 dot) grid, as  $2 * (5)\text{dots/row} * (5 - 1) = 40$ . Then, each successive level has one less move. So looking two moves ahead from the start yields 39 moves for each of the 40 first moves, or 39 branches off of each of the 40 original branches. As the levels progress, there are 38, then 37, then 36, and so on. Looking all 40 moves ahead, the total number of possible board layouts is  $40!$ , which is equal to 815,915,283,247,897,734,345,611,269,596,115,894,272,000,000,000 possible final board states. The computer would have to consider all of these possibilities if it were to look through the complete set of moves all the way to the end of the game. However, this is not practical, simply because of the sheer computational power and time that would need to be allocated for this. Due to this limitation, our algorithm is set up to only look 7 levels deep. Although, even at 7 moves ahead, the program still ran quite slowly.

In order to reduce the computation time of each move, we implemented a technique called alpha-beta pruning into our code. Alpha-beta pruning is an algorithm used in the minimax algorithm that reduces the number of possible moves that the computer would have to search through to find the best possible move.<sup>1</sup> It does this by eliminating, or “pruning”, branches of the tree that would anyways be discarded by the minimax algorithm, and therefore are unnecessary to go through in the first place.

Alpha-beta pruning utilizes a maximizer and a minimizer, where the maximizer is the player attempting to attain the maximum score (generally the player who goes first), while the minimizer is the player that attempts to lower the maximizer’s best possible score (typically the player who goes second). Alpha-beta pruning gets its name from the two parts it considers:

- Alpha, which is the best board (highest score) the maximizer can decidedly achieve by following the tree to its maximum depth.
- Beta, which is the best board (lowest score) the minimizer can decidedly achieve by following the tree up to its maximum depth.

At each minimizer level of the tree, the computer will compare the alpha value of the level above to the current beta value. When it knows this beta value, or the lowest score that the minimizer found, it can confidently say that the final score that the minimizer will pick will be equal to or lower than this value. This means that we can compare what we have found as the highest value for the maximizer, the alpha value, to this current beta value. If this beta value is less than the alpha we can confidently say that the maximizer will never go down the path with this beta value and we can stop exploring this section of the tree.<sup>1</sup> This can reduce the amount of cases we have to look through later which will be helpful in optimizing our code.

**Figure 4: Alpha-Beta Pruning Example**

As shown in Figure 4, alpha-beta pruning reduces the computations necessary to play out the optimal moves since there is no longer a need to explore every leaf of the tree. The program can instead prioritize the most important ones and spend its time calculating the most likely paths.

### III. Functionality

#### A. Setup

To create our code base we decided to use the MVC or Model View Controller coding schema, which consists of using 3 different classes, the Model, View, and Controller. The Model deals with the fundamental data of the code (the different values in the board) and contains functions needed to find and manipulate the data. The View class handles the display of the data, in our case working with the pygame library to output the board to the screen as well as printing out a text version of it for later analysis. The third class, the Controller handles the logic of the game and controls the computer's moves.

#### B. Functions

##### 1. Model

The Model class is the section that contains the game data as well as managing functions to retrieve and change that data. Since any game board size can be used for Dots and Boxes, we chose to use a 5x5 dot grid (16 boxes) as it was big enough to simulate some of the nuances of the game without causing issues with having huge numbers of moves. We used a 2D array, shown in Figure 5, to represent the rows and columns of the board and set it equal to the `self.data` variable. For the empty board, we used "\*" to represent dots, "!" to represent unfilled lines, and "0" to represent the inside of unclaimed boxes, all of which are unclaimed at the beginning of the game. As the game goes on, we continue to update our model to reflect what lines have been made and which players have claimed each box. When lines are made, the exclamation points are turned into "/" to represent horizontal and vertical lines, and when boxes are made, the "0" is changed to either "c" or "h" (computer or human) depending on which player owns the box.

```

    "o = EmptyBox
self.data=[[["*", "!", "*", "!", "*", "!", "*", "!", "*"], 
           ["!", "0", "!", "0", "!", "0", "!", "0", "!"], 
           ["*", "!", "*", "!", "*", "!", "*", "!", "*"], 
           ["!", "0", "!", "0", "!", "0", "!", "0", "!"], 
           ["*", "!", "*", "!", "*", "!", "*", "!", "*"], 
           ["!", "0", "!", "0", "!", "0", "!", "0", "!"], 
           ["*", "!", "*", "!", "*", "!", "*", "!", "*"], 
           ["!", "0", "!", "0", "!", "0", "!", "0", "!"], 
           ["*", "!", "*", "!", "*", "!", "*", "!", "*"]]] 
self.spacesleft=40

```

**Figure 5: Model Code**

## 2. playGame

The game starts by asking the user how many players (0-2) they want to play with and subsequently calls the *playGame* function using that value. As shown in Figure 6, If the player inputs “0” then the computer will play against itself. If the player inputs “1” the player will play against the computer, and if the player inputs “2” then two humans can play against each other. Figure 6 exhibits how the *playGame* function draws the board onto the screen, explained later in the View section, and then assigns the variables *currentplayer* and *nextplayer* to either “h1”, “h2”, “c1”, or “c2”, based on who is playing the game. “H” and “c” indicate humans and computers, while the numbers “1” and “2” represent which player goes first and second. The game will start playing and continue to play until no empty spaces are left on the board. When that happens, the game is over and the program ends.

```

# Function to play the game
def playGame(self,playernum):
    # Start displaying the game board using Pygame
    self.view.startDisplay()

    self.view.drawBoard()
    #determines the player types for both players as well as the
    corresponding values
    if playernum==1:
        currentplayer="h1"
        nextplayer="c2"
    elif playernum==2:
        currentplayer="h1"
        nextplayer="h2"
    else:
        currentplayer="c1"
        nextplayer="c2"
    # Play turns until there are no more empty spaces on the game board
    while self.model.spacesleft>0:
        playAnother=self.playTurn(currentplayer,nextplayer)
        self.view.drawBoard()
        if not playAnother:
            currentplayer,nextplayer=nextplayer,currentplayer
            self.model.spacesleft=self.model.countSpots(self.model.data)
            self.file.close()

```

**Figure 6: *playGame* Function**

### 3. playTurn

Figure 7 is the *playTurn* function, which is enacted when a player takes a turn. A turn consists of the complete set of moves that a player makes at once. This can range from a singular move to multiple moves, since if a player uses their move to complete a box, they must make another move. The function starts by detecting when each player makes a move and updating the model to represent that move. It then checks to see if the move that the player made completed a box on the board. If it did, the player is allowed to take another turn. If it didn't, the updated game board is drawn onto the screen and the next player takes their turn. The *playTurn* function is called inside of the *playGame* function until there are no more empty spaces left on the game board.

```
def playTurn(self,player,opp):
    if player[0] == "h":
        redo=True
        while redo:
            # Wait for a mouse click as input from the player
            pos=self.waitForResponse(player,opp)
            takeAnotherTurn=False
            takeAnotherTurn2=False
            for i in range (len(pos[0])):
                time.sleep(0.1)
                x=pos[0][i][1]
                y=pos[0][i][0]
            # Update the game board based on player's input
            if self.model.data[y][x]=="!":
                self.file.write(player+":"+str(x)+"-"+str(y)+"\n")
                self.model.updateValue(y,x,'/',self.model.data)
                takeAnotherTurn=False
                takeAnotherTurn2=False
                if y==0:
                    takeAnotherTurn=self.checkBox(1,x,self.model.data, player)
                elif y==8:
                    takeAnotherTurn=self.checkBox(7,x,self.model.data, player)
                elif y==2 or y==4 or y==6:
                    takeAnotherTurn=self.checkBox(y-1,x, self.model.data,player)
                    takeAnotherTurn2=self.checkBox(y+1,x,self.model.data, player)
                if x==0:
                    takeAnotherTurn=self.checkBox(y,1, self.model.data,player)
                elif x==8:
                    takeAnotherTurn=self.checkBox(y,7, self.model.data,player)
                elif x==2 or x==4 or x==6:
                    takeAnotherTurn=self.checkBox(y,x-1,self.model.data, player)
                    takeAnotherTurn2=self.checkBox(y,x+1,self.model.data, player)
            # Draw the updated game board
            self.view.drawBoard()
            redo=False
        else:
            print("ya stupid")
```

Figure 7: *playTurn* Function

#### 4. flipBoard

As demonstrated in Figure 8, the *flipBoard* function exchanges the roles of the two players. It does this by iterating through the board, checking who owns each box on the board, and then switching ownership to the opposing player. For example, if the computer owns 5 boxes, those boxes will be flipped so that the human owns them. In the same way, the computer will then take ownership of the human's boxes. The purpose of this function is to be able to look ahead at the possible moves of the opponent from the opponent's point of view. That way, the computer can see what move its opponent would make if it was playing optimally and respond accordingly. This function is important for the *generateTree* function: every time the computer goes down another level in the decision tree, the board is flipped so that the computer can look at moves from the correct point of view.

```
#exchanges the two side's designations
def flipBoard(self,chars,board):
    for i in range (len(self.model.data)):
        for j in range (len(self.model.data[0])):
            if board[i][j]==chars[0]:
                board[i][j]=chars[1]
            elif board[i][j]==chars[1]:
                board[i][j]=chars[0]
    return board
```

Figure 8: *flipBoard* Function

#### 5. getEmptySpaces

Figure 9 shows the *getEmptySpaces* function which returns a list of all legal turns that the computer can take. These turns could include multiple individual moves. It first looks for lines on the board that haven't been made yet and then checks to see if making a line there will complete a box. Once it has identified a move that can finish a box it then repeats the process to get a set of all moves that complete boxes. Then, since every time you complete a box you get to make another move, it then makes one copy of this set and adds each move that does not complete a chain to the end of a copy. This provides all of the turns that complete the maximum amount of boxes and then plays its next move. Figure 9 contains the code for this function.

However we were not done yet, since the set of all legal moves also include those that do not fill all the possible boxes, but we discovered that strategically the only other set we had to consider was the one that left the last 2 boxes unfilled, but filling in the 3rd side of one, also referred to earlier as a double-cross. This was complicated by the fact that chains that had 3-sided squares on each end couldn't be double-crossed like usual and had to sacrifice 4 boxes instead of just 2, and also the fact that double-crossing was not always possible in certain board states. However we were eventually able to correctly identify the full set of legal moves. This function is important so we can consider all of the relevant moves for each level in the tree.

```

def getEmptySpaces(self,board,count,player):
    chains=[]
    legals=[]
    Bchain=True
    newboard = copy.deepcopy(board)
    while Bchain: #Bchain makes sure it only adds one legal move at a time
        Bchain=False
        for y in range(len(newboard)):
            for x in range(len(newboard[0])):
                if (not Bchain) and newboard[y][x] == "!" and self.checkIfMakesBox(y,x,newboard):
                    newboard=self.playMove([(y,x)],newboard,player)
                    Bchain=True
                    val=(y,x)
                    chains.append(val)
    checker=False
    for y in range(len(newboard)):
        for x in range(len(newboard[0])):
            if newboard[y][x] == "!":
                checker=True
                newchains=copy.deepcopy(chains)
                newchains.append((y,x))
                legals.append(newchains)
    if not checker:
        legals.append(chains)

```

Figure 9: *getEmptySpaces* Function

## 6. *scrapeMoves*

To manage the huge volume of moves that we have collected through the *getEmptySpaces* function we created the *scrapeMoves* function, depicted in Figure 11. This function utilizes the fact that while different moves are initially distinct, they can sometimes have the same result after the next player plays their turn. Specifically when playing moves into a chain, any move that starts a cascade from a point in the chain is functionally identical:

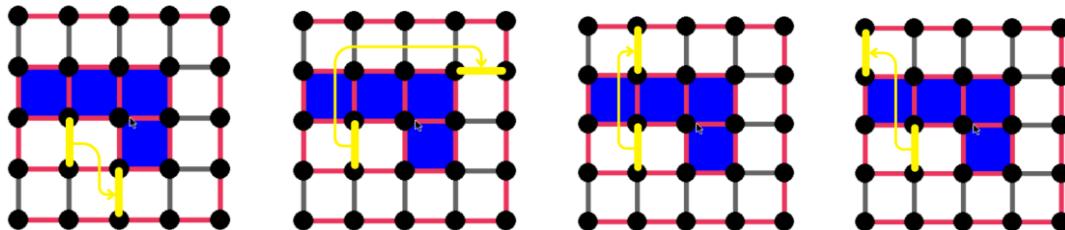


Figure 10: Scraping Moves Example

Since all of these moves shown in Figure 10 result in the opponent winning the entire outer cascade, we can delete all but one from our move list. To do this we first make a new list of the empty sides of each box with the *findDirections* function. We then use this list to run the *followChains* function, which picks a box and follows its empty sides until we find a box that doesn't have 2 sides completed and thus isn't part of a chain. We then take this list of boxes, turn it into a list of lines with the *interpretChain* function and remove any move that ends in any except for the first of those lines. This lets us not have to consider identical board states multiple times.

```

def scrapeMoves(self,moves,board,bool=False):
    interpretedchain=self.followChains(board,self.findDirections(board))
    movescopy2=copy.deepcopy(moves)
    for i in range(len(movescopy2)):
        for chainedmoves in interpretedchain:
            copying=copy.deepcopy(chainedmoves)
            temp=copying.pop(-1)
            try:
                if movescopy2[i][-1] in copying:
                    moves.remove(movescopy2[i])
            except:
                pass
    return moves

```

**Figure 11: scrapeMoves Function**

## 7. scoreBoard

The *scoreBoard* function is used to evaluate the favorability of a game board for a given player, computer or human. A game board is considered more favorable for a player if the player has claimed more boxes than their opponent. Our *scoreBoard* function iterates through the game board, which is a 2-dimensional array, and increments a counter variable when it finds a box that is owned by the player, and decrements the counter when it reaches a box controlled by the opponent. Essentially, it finds the difference between the number of boxes that the player controls and the opponent controls in that particular version of the game board. The counter will be positive if the player has more boxes than their opponent, zero if they have the same number of boxes, and negative if the player is at a disadvantage with less number of boxes than their opponent. These are the positive and negative 1 in Figure 12. This means that the computer will always view claiming more boxes as a higher score and thus a better move. This will help the computer win more often as it will always be trying to create a winning board. This function is crucial when comparing favorabilities of possible game boards to determine the best case scenario for the player later on.

```

#scores the board, currently based on how many
boxes each player controls
def scoreBoard(self,board,player,opp):
    counter=0
    for i in range(len(board)):
        for j in range(len(board[0])):
            if (i+j)%2==0 and i%2==1:
                if board[i][j]==player:
                    counter+=1
                elif board[i][j]==opp:
                    counter-=1
    return counter

```

**Figure 12: scoreBoard Function**

## 8. generateTree

The *generateTree* recursive function looks ahead at all of the possible game boards that would be the result of any of the possible moves that can be made by both the opponent and the player. It begins by creating a deep copy of the current board so that it can consider the scores of all of the possible future boards without altering the current board. The function takes in a *targetDepth*, which represents the number of

moves ahead, or levels deep in the decision tree, that it will examine. The base case, located near the top of Figure 13, is when the target depth is 0, so it just returns the best move on the lowest level it's looking at. If it's not at the lowest level, it gets a list of all the legal moves that should be considered through the `getEmptySpaces` function as described above. Once it has done this it procedurally plays each move and reruns the `generateTree` function at one level lower of `targetDepth`. This means that the tree is recursively generated and every sequence of moves is considered. Then after generating each tree, it returns the best move it has found as well as the score it is giving that move. Then the next level up takes the values that return from the mini-tree of each move and finds the best move by comparing those values. This gives a very full picture of the set of possible moves, but is also an incredibly high number. To combat this we implemented alpha-beta pruning. To do this we simply used a variable to record the best value found a level above where a mini-tree is getting created and if that value is either less or greater, depending if we are on a maximizing or minimizing level, than the move we are currently inspecting, we simply return the current move and “prune” off this branch. We can do this since when two optimal players are playing, this branch will never be played and thus has no need to be calculated.

```
## the look ahead function
## returns the optimal value and move that the maximizer can obtain
def generateTree(self, board, targetDepth, player, opp, alpha):
    theboard=copy.deepcopy(board)
    if targetDepth == 0:
        if self.countBlanks(theboard)>0:
            move=self.playBest(theboard, player, opp, False, alpha)
            return move
        else:
            move=[[[],0]],self.scoreBoard(theboard,player,opp)
            return move
    else:
        if self.countBlanks(theboard)>0:
            leaves = self.getEmptySpaces(theboard,0,player)
            leaves = self.scrapeMoves(leaves,theboard,True)
            random.shuffle(leaves)
            score = -17
            maxmove=[[10,10]]
            for i in range(len(leaves)):
                newBoard = self.flipBoard([player,opp],copy.deepcopy(theboard))
                newBoardCopy=self.playMove(leaves[i],newBoard,player)
                temp=[leaves[i]]
                temp.append(self.generateTree(newBoardCopy,targetDepth - 1, player, opp, score)[1])
                leaves[i]=temp
                if leaves[i][1] >= -alpha:
                    print(leaves,targetDepth)
                    return[leaves[i][0],-leaves[i][1]]
            if(leaves[i][1] > score):
                score = leaves[i][1]
                maxmove=leaves[i][0]
            print(leaves,targetDepth)
            return [maxmove,-score]
```

Figure 13: `generateTree` Function

## C. GUI

Early on in our testing process, we discovered that working solely with a coordinate based input system was slow for testing and hard to interpret. To make testing and playing the game easier, we decided to implement a graphical user interface (GUI) into our code. To do this we used the library pygame, which allows for drawing simple shapes to a screen and interpreting keyboard commands such as keystrokes and clicks. We used this to create a visual representation of the game board that can have moves inputted by

clicks. The steps to this process were: drawing a blank board, updating the board so it displays game data, and being able to receive moves via clicks.

First to draw the blank board we use the *drawBlankBoard* function, which works with a variable called *boxsize* to define how spread apart the lines and dots should be, then using some for-loops to draw all the needed lines and dots.

Then the *drawBoard* function fills in the lines and boxes to match the data from the board. It loops through all the spaces on the board, and in each section, after identifying which spaces it cares about through the use of modular arithmetic, checks all of the relevant values to see if a new line or box needs to be drawn. It then draws the appropriate boxes and lines, with boxes being drawn first so that when it has finished the lines are not cut off. Finally since the lines are drawn as long rectangles, we need to re-draw the dots so that they appear over the lines, which we do just as in the *drawBlankBoard* function.

The last aspect of the GUI is the *waitForResponse* function, which will intercept and interpret player clicks. When an event happens in pygame it is given a location and a type, so we simply search for the "MOUSEBUTTONDOWN" event and take its coordinates. We wanted to make the interface intuitive to use, so instead of mandating that the player clicks exactly on the line they want to pick, we create a diamond-shaped "hitbox" around each line. The resulting grid of diagonal lines is shown in Figure 14. To do this we used a set of lines and equations to find the line that was closest to the player's click input. This means that the click has a good amount of room for error making the interface quicker and less annoying to use.

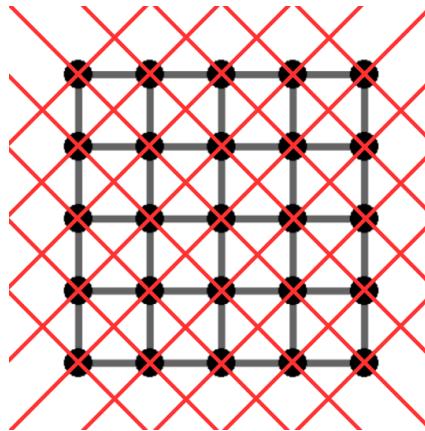


Figure 14: Pygame Interfacing

## IV. Testing Our Code

### A. Successes

Though our team quickly became concentrated on trying to debug code and fix setbacks, this did not take away from our successes. First, creating a fully functional game of dots and boxes for two players, both in the console and then using the interactive graphical user interface, was an encouraging step forward in this project. Additionally, being able to implement several aspects of advanced computing strategy, such as the minimax tree, alpha-beta pruning and accounting for double-crosses has been particularly successful. While the computer may not be able to win every time, having a high success rate of wins while also having reasonable computational time for the computer moves was a goal we were very happy to achieve.

## B. Setbacks

Computation time quickly became a major obstacle for this project. As was iterated in our explanation of alpha-beta pruning, a 4 by 4 square game of Dots and Boxes has  $40!$  Possible ending boards. Our early code attempted to play through every possible move combination for each level (turn) ahead it looked. Trying to reach 3 or more levels deep was not sustainable as we did not have the computational power to do so in a practical time slot. The amount of time needed for the computer to complete a turn increased exponentially with each additional level deep. This problem worsened with the added implementation of our double-crosses strategy and alpha-beta pruning. We mitigated this problem with the inclusion of our scrapeMoves function along with switching to lists of moves instead of sets, which were very harmful to the processing time. Once this was done, alpha-beta pruning helped to consider only the best moves for the current player to make.

## V. Trade-offs

### A. Java versus Python

The two languages that our team considered for this project were Java and Python. Java, being a compiled language, requires the code to be compiled prior to running. Hence, any errors are caught before execution of the code. This allows it to run faster and be easier to debug than Python. Python is an interpreted language, where code is executed directly without a preceding compilation step. Being an interpreted language, Python's syntax is less lengthy and complex than Java's syntax. Therefore, it is more beginner friendly with less effort needed to write and read code. Ultimately, our team made the decision to code in Python. We chose to prioritize readability of the language at the expense of runtime. Using Python also allowed us to avoid excessive syntax errors due to it being a dynamic typed language.

### B. Computational time

One of the tradeoffs we had to make while coding was trading “perfection” for practicality. Every time the computer tries to go a level deeper in the minimax tree, the computational time increases exponentially. In theory, the best minimax algorithms can look to the end of the game and score the board. For our Dots and Boxes board, there are 40 possible moves at the beginning, so the perfect algorithm would be able to look 40 moves ahead. But because this would take a drastic amount of time and power that we do not have, the deepest level our algorithm can reach is level 7. Also, the computer takes even more time to consider double-crosses. Combining deeper levels with double-crossing results in a runtime of about 10-15 seconds at most. Since we go down seven levels and we start counting at level 0, there are 8 total levels that our computer considers and the computational time of the program corresponds to  $O(n^8)$  at most.

## VI. Data Analysis

As we progressed in research of Dots and Boxes and in the creation of the game’s code, the direction of this project started to become clear.

- . The key focus of this project turned towards these questions:
  1. Does double-crossing have a noticeable impact?
  2. Is having either the first or second move an advantage? Does it matter?
  3. Can our program beat a human with little to no prior experience with Dots and Boxes?

The following sections will show our analysis of the data we collected to answer these questions.

### A. No-DC Code vs. Optimal Human

Throughout this project, an “optimal human” played against both versions of our code: one that can not double-cross its opponents (No-DC Code), and one that does double-cross its opponents (DC Code).

The first version of the game implemented a minimax algorithm that could not double-cross its opponents, hence the name No-DC (no double-cross). The inability to double-cross severely hindered this version, allowing the human player to score more boxes by exploiting the algorithm’s weaknesses.

Table 2 shows that the mean human score with the optimal human player playing first was 8.467 boxes, with a standard deviation of  $\pm 0.6399$  boxes. The distribution of the data is skewed to the left, as the majority of the scores (60%) were draws where the human player scored 8 points. There was no score less than 8, as the human player only drew or won the games.

The mean human score with the optimal human player playing second was 9.733 boxes, with a standard deviation of  $\pm 0.8837$  boxes. The distribution of the data is skewed to the right, as the majority of the scores (60%) were wins where the human player scored 10 or more points. As shown in Table 1, there was no game in which the optimal human player scored less than 8, as the human player only drew or won the games.

When the optimal human went second against the No-DC Code, the human scored higher. This is shown in Tables 1, 2 and 3. In all future set-ups, the data show the opposite trend. This is because the No-DC Code is incapable of double-crossing. The optimal human player can set up a scenario in which the No-DC Code either misses an opportunity to double-cross, or the optimal human player can double-cross the No-DC Code. The entire point of double-crossing is to switch the parity of the game; the player that has to start the first long chain loses the game, assuming the opponent knows how to double-cross. Normally, the first player attempts to create opportunities to double-cross its opponent. However, the No-DC Code is incapable of double-crossing. Table 3 shows how the optimally-playing human exploited this aspect of the No-DC Code, winning or drawing every single game against it. Importantly, neither going first nor second enabled the No-DC Code to defeat the optimal human player in any single game.

**Table 1: Raw Data for No-DC Code vs. Optimal Human**

Game Number	Optimal Human Score (Going First)	No-DC Code (Going Second)	Game Number	No-DC Code (Going First)	Optimal Human Score (Going Second)
1	9	7	1	5	11
2	9	7	2	6	10
3	8	8	3	7	9
4	8	8	4	7	9
5	8	8	5	7	9
6	8	8	6	8	8
7	8	8	7	6	10
8	9	7	8	7	9
9	9	7	9	5	11
10	10	6	10	6	10
11	8	8	11	6	10
12	8	8	12	6	10
13	8	8	13	5	11
14	9	7	14	7	9
15	8	8	15	6	10

**Table 2: Averages and Standard Deviations for No-DC Code vs. Optimal Human**

	Human Score	No-DC Code Comp. Score	Human Moves Number	No-DC Code Comp. Moves Number
Averages & Standard Deviations:				
Human Goes First, Averages:	8.467	7.533	21.2	18.8
Human Goes First, St. Devs.:	0.6399	0.6399	0.6761	0.6761
Computer Goes First, Averages:	9.733	6.267	21.93	18.067
Computer Goes First, St. Devs.:	0.8837	0.8837	0.7988	0.7988

**Table 3: Outcomes Percentages for No-DC Code vs. Optimal Human**

	Win Percentage:	Draw Percentage:	Loss Percentage:
Human Goes First, Human:	40.00	60.00	0.00
Human goes Second, Human:	93.33	6.67	0.00
Computer Goes First, Computer:	0.00	60.00	40.00
Computer Goes Second, Computer:	0.00	6.67	93.33

## B. No-DC Code vs. DC Code

Our objective was to create a program that could both look-ahead and double-cross its opponent. The new version of the code could double-cross (DC Code) and implement look-ahead in stages: for the first 15 moves, the algorithm looks 3 levels deep. Then, for the next five moves (moves 16-20), it looks 5 levels ahead. For the remaining 20 moves, it looks 7 levels ahead.

The series of games between No-DC Code vs. DC Code serves both as a potential confirmation of the results of the previous set-up (No-DC Code vs. Optimal Human) and a potential isolation of the reason why No-DC Code loses frequently to a human player. If the DC Code performs better than the No-DC Code, then implementing double-crossing has a positive effect on the success of the algorithm. If the No-DC Code defeats the DC Code, then double-crossing would have no overall effect on program success. The raw data for these trials are shown in Table 4.

From Table 5, one may notice that when DC Code moves first, the mean number of boxes scored is 9.500, while the No-DC Code scores 6.5 boxes on average. The standard deviation for this data set is  $\pm 2.0138$  boxes. When the No-DC Code goes first, it scores 9.000 boxes on average, while the DC Code has a mean score of 7.0 boxes. DC Code scores slightly higher than the No-DC Code overall, yet Table 6 demonstrates that DC Code has a much greater win percentage going first (60% to 20%) and equal win percentage to the No-DC code when DC Code goes second (30% each). Even though the average scores are similar for either algorithm when making the first move, DC Code wins more often across both sets overall.

It's important to note that the algorithm treats all losses as the same. A loss 7-9, from the DC Code's perspective, is equally as terrible an outcome as a loss 0-16. This is why, when the DC Code lost, it lost by a margin relatively larger than in its wins over the No-DC Code. Table 4 contains the results of each game and shows that particularly, the DC Code's three losses when going second were 4-12, 1-15, and 0-16, while the No-DC Code's 6 losses had final scores of 4-12, 6-10, or 7-9. The algorithm rates all lost boards (boards with no possible winning outcome) with a value approximately equal to  $-3 \times 10^{15}$ , which has a much greater magnitude than any of the other board scoring techniques. This ensures that the DC Code avoids all lost boards at all costs.

The standard deviation of the second set (where DC Code goes second) is more than double the standard deviation of the first set. There is a much greater range of scores. When the No-DC Code played the first move, the scoring was inconsistent, with a data range of 12 (from 4 boxes at the least to 16 boxes at the greatest). When DC Code played the first move, the scoring was more consistent, with a data range of 5 (from 7 boxes at the least to 12 boxes at the greatest). This discrepancy is caused by the scoring algorithm in DC Code: the algorithm calculates all losing boards as the same. From the algorithm's perspective, losing 9-7 is equally bad as losing 16-0. The fact that DC Code never scored less than 7 boxes when moving first demonstrates its consistent control over the No-DC Code.

As shown in Table 6, the DC Code did not have a losing record going second and did have a significantly winning record going first. This shows that implementing double-crossing is a crucial element of creating a well-performing Dots and Boxes algorithm.

**Table 4: Raw Data for No-DC Code vs. DC Code**

Game Number:	No-DC Code (goes first) Score:	DC Code (goes second) Score:	Game Number:	DC Code (goes first) Score:	No-DC Code (goes second) Score:
1	6	10	1	12	4
2	12	4	2	8	8
3	8	8	3	12	4
4	4	12	4	12	4
5	8	8	5	10	6
6	5	11	6	7	9
7	15	1	7	8	8
8	8	8	8	10	6
9	8	8	9	9	7
10	16	0	10	7	9

**Table 5: Averages and Standard Deviations for No-DC Code vs. DC Code**

Averages & Standard Deviations:		
	No-DC Code Score:	DC Code Score:
DC Code Goes First, Averages:	6.5000	9.5000
DC Code Goes First, St. Devs.:	2.0138	2.0138
No-DC Code Goes First, Averages:	9.0000	7.0000
No-DC Code Goes First, St. Devs.:	4.0552	4.0552

**Table 6: Outcomes Percentages for No-DC Code vs. DC Code**

Win, Draw, & Loss Percentages:			
	Win Percentage:	Draw Percentage:	Loss Percentage:
DC Code Goes First, DC Code:	60.00	20.00	20.00
DC Code goes Second, DC Code:	30.00	40.00	30.00
No-DC Code Goes First, No-DC Code:	30.00	40.00	30.00
No-DC Code Goes Second, No-DC Code:	20.00	20.00	60.00

### C. DC Code vs. DC Code

The next key question this project sought to answer was whether going first or second yielded any advantage, or whether it did not have a noticeable effect on the outcomes. To answer this question, the next series of games was played between two identical copies of the DC Code.

Because the DC Codes are identical to each other, there needed to be a distinction between them. C1 represents the DC Code that went first, and C2 represents the DC Code that went second. This allowed the algorithm to play itself. If the average scores were identical, then there would be no advantage to moving first or second. If the average score of either C1 or C2 was greater than the other, then it would indicate a noticeable advantage for making either the first or second move. At this point in the project, it was not yet known whether going first or second had any impact on the results, as the optimal human did better against the No-DC Code going second, while the first player in the No-DC Code vs. DC Code series of games performed better.

Table 8 shows the significant difference between the two options: C1 had a mean score of 10.0667 boxes, nearly double that of C2 (5.9333 boxes). From Table 7, C1 had the most skewed victories over C2, winning with scores such as 13-3, 14-2, and even 16-0. The dominance of C1 is expressed in Table 9: winning a majority of the games and drawing a third of them, C1 decisively defeated C2. Therefore, going first is advantageous against opponents capable of double-crossing.

There was one significant setback to note: the DC Code was set to run against itself at each possible level of recursion. This would have added an extra element to the project, assessing how strength of the program changes (if at all) as the level to which it looks ahead increases. However, a computer system refresh deleted the games saved on the computer. This pivoted the direction of our research: testing the deepest level of recursion against itself. After all, the DC Code is based on looking as deep as it can in a reasonable amount of time; this is the most reliable test to prove that the player going first has an advantage.

**Table 7: Raw Data for DC Code vs. DC Code**

GAME Number:	C1 (goes first) Score:	C2 (goes second) Score:
1	8	8
2	11	5
3	9	7
4	10	6
5	16	0
6	13	3
7	8	8
8	4	12
9	8	8
10	7	9
11	8	8
12	14	2
13	14	2
14	8	8
15	13	3

**Table 8: Averages and Standard Deviations for DC Code vs. DC Code**

Averages & Standard Deviations:		
	DC Code (Going First) Score	DC Code (Going Second) Score
DC Code Goes First, Averages:	10.0667	5.9333
DC Code Goes First, St. Devs.:	3.3051	3.3051

**Table 9: Outcomes Percentages for DC Code vs. DC Code**

Win, Draw, & Loss Percentages:			
	Win Percentage:	Draw Percentage:	Loss Percentage:
DC Code Goes First C1:	53.33	33.33	13.33
DC Code goes Second, C2:	13.33	33.33	53.33

## D. Average Human vs. DC Code

Implementing strategies, testing different versions of the algorithm, and analyzing the advantages of move order all enhance the code. Yet, there was still a necessary test to see if these could all come together to answer the final key question: Can our best algorithm beat a human player? The following series of games involved an average human player with little prior experience playing Dots and Boxes facing the DC Code.

Inferring from the previous two series, both players should perform better when going first than when going second. However, the DC Code should be able to definitively defeat the average human. An interesting situation arises when balancing the two ideas: can the average human do notably better going first while still losing to the DC Code?

Table 10 displays the scores of each game played by the average human against the DC Code. The average human was able to score 10-6 and 12-4 when going first. All other games were either losses to the DC Code or draws. The mean score for the average human when going first was 6.200 boxes, and the mean was 4.100 boxes when going second, as shown in Table 11. This notable discrepancy indicates that going first is still advantageous, even to a player who does not perform as well as the DC Code. Table 12 visibly represents the DC Code outperforming the average human, winning 100% of all the games when going first and 60% of the game when going second. These large margins of victory vindicate the DC Code's successful implementation of the strategies and look-ahead functions.

**Table 10: Raw Data for Average Human vs. DC Code**

Game Number:	Average Human (goes first) Score:	DC Code (goes second) Score:	Game Number:	DC Code (goes first) Score:	Average Human (goes second) Score:
1	3	13	1	12	4
2	8	8	2	9	7
3	2	14	3	15	1
4	12	4	4	15	1
5	4	12	5	11	5
6	5	11	6	11	5
7	8	8	7	12	4
8	3	13	8	13	3
9	10	6	9	10	6
10	7	9	10	11	5

**Table 11: Averages and Standard Deviations for Average Human vs. DC Code**

Averages & Standard Deviations:		
	Average Human Score	Computer Score
Average Human Goes First, Averages:	6.2000	9.8000
Average Human Goes First, St. Devs.:	3.3267	3.3267
DC Code Goes First, Averages:	4.1000	11.9000
DC Code Goes First, St. Devs.:	1.9692	1.9692

**Table 12: Outcomes Percentages for Average Human vs. DC Code**

Win, Draw, & Loss Percentages:			
	Win Percentage:	Draw Percentage:	Loss Percentage:
Average Human Goes First, Human:	20.00	20.00	60.00
Average Human Goes Second, Human:	0.00	0.00	100.00
Computer Goes First, Computer:	100.00	0.00	0.00
Computer Goes Second, Computer:	60.00	20.00	20.00

The DC Code provides a challenge on a user-friendly interface for human players. We achieved the goal of creating a Dots and Boxes game to both entertain and educate its opponents. A new question arose: how will the DC Code fare against an optimal human player?

## E. Optimal Human vs. DC Code

The optimal human is a person who has practiced Dots and Boxes by playing hundreds of games and learning about the strategies behind the game. This person was fully aware of double-crosses and other structures that could help gain a winning position on the board. Also, the optimal human was able to quickly calculate many levels deep, especially once all moves would start a cascade. This is because the cascades can be simplified into one set of identical moves (any move in a cascade allows it to be taken). Similar to the `scrapeMoves` function in the code, the optimal human player treats all moves in a given cascade as the same move (except for double-crosses), decreasing the brainpower necessary to play optimally.

Table 13 shows the results of the games played between the optimal human and the DC Code. The majority of the games in each set were either draws (8-8) or close games with scores of 7-9 or 9-7. Some of the games in which the DC Code played first ended in a 13-3 victory for the code. These games involved exceedingly long chains (typically 7 or more boxes long) and double-crossing by the DC Code forced the opponent to give up the long chain. In the majority of games, however, the optimal human managed to section off smaller chains and avoid these drastic losses.

Table 15 shows the overall results: in both sets, the optimal human won a majority of the time (55%). When going first, the optimal human only lost 5% of the games, but the DC Code held the human to 40% draws. When the optimal human played the second move, the DC Code won a larger share of the games (35%). While the optimal human performed better overall, the DC Code performed better than the No-DC Code did against the optimal human player. The DC Code was capable of defeating the optimal human player, providing the greatest performance when going first.

Observing the score margins shown in Table 14, one can tell that the average game was much closer between the DC Code and optimal human than for any other series of tests. The average human scores were 8.850 boxes and 8.050 boxes when going first and second, respectively. The 8.050 boxes is barely above the average score for the DC Code going first (7.950 boxes). The DC Code creates opportunities to win even against experienced opponents.

**Table 13: Raw Data for Optimal Human vs. DC Code**

Game Number:	Optimal Human (goes first) Score:	DC Code (goes second) Score:	Game Number:	DC Code (goes first) Score:	Optimal Human (goes second) Score:
1	9	7	1	7	9
2	8	8	2	7	9
3	9	7	3	10	6
4	8	8	4	9	7
5	8	8	5	7	9
6	8	8	6	8	8
7	8	8	7	8	8
8	8	8	8	7	9
9	11	5	9	7	9
10	10	6	10	9	7
11	8	8	11	9	7
12	5	11	12	7	9
13	9	7	13	7	9
14	8	8	14	5	11
15	11	5	15	13	3
16	9	7	16	10	6
17	11	5	17	13	3
18	9	7	18	7	9
19	10	6	19	5	11
20	10	6	20	4	12

**Table 14: Averages and Standard Deviations for Optimal Human vs. DC Code**

Averages & Standard Deviations:		
	Optimal Human Score	DC Code Score
Optimal Human Goes First, Averages:	8.8500	7.1500
Optimal Human Goes First, St. Devs.:	1.4244	1.4244
DC Code Goes First, Averages:	8.0500	7.9500
DC Code Goes First, St. Devs.:	2.3278	2.3278

**Table 15: Outcomes Percentages for Optimal Human vs. DC Code**

Win, Draw, & Loss Percentages:			
	Win Percentage:	Draw Percentage:	Loss Percentage:
Human Goes First, Human:	55.00	40.00	5.00
Human goes Second, Human:	55.00	10.00	35.00
Computer Goes First, Computer:	5.00	40.00	55.00
Computer Goes Second, Computer:	35.00	10.00	55.00

The data provides answers to the key questions we sought to answer: First, double-crossing is beneficial and does have an impact on performance. Second, having the first move is advantageous. Third, the DC Code is capable of decisively defeating an average human, even challenging an experienced player who implemented optimal strategies.

## VII. Conclusion

### A. Discussion

The research done into Dots and Boxes proved to be enlightening in several different game aspects. What seemed to be a simple pen-and-paper game was capable of being a source of complex, in-depth analysis. Our research showed us how mathematical strategy and game theory could be applied to seemingly simple situations in order to learn about them. We were first able to analyze games of Dots and Boxes and found that there were consistent trends and strategies present in wins. Strategies such as double-crossing and avoiding initiating cascades proved to be crucial to our strategy implemented in our DC Code. In addition to finding various strategies to add into the algorithm, we were able to implement a minimax decision tree

and algorithm to sort through and choose the best moves. However, this did come with some restrictions, as we discovered some limitations of the minimax tree. To partially solve this issue, we implemented Alpha-Beta pruning. We also discovered that there was a need to hard code specific strategies such as double crossing because we could not look down the entire minimax tree. This taught us about how to choose our coding priorities, and how a computationally-adjusted solution can solve a problem that a purely computational one struggles with. Lastly, we ran data analysis on our code and compared it to four opponents: No-DC Code, DC Code (itself), an average human, and an optimal human. Comparing with outside sources can be very valuable in evaluating code and algorithms. This project allowed an in-depth exploration of turn-based game strategy and the use of intuitive, computational solutions that balance well with optimal strategy.

## B. Future Improvements

There are many aspects of this program that we could improve in the future. For example, there are some visual aspects that we could enhance. In a game of dots and boxes, it is more visually appealing to color code the lines drawn on the board for each player. This draws the eye to the most recent move and who the move was made by. The reason we did not incorporate this into our existing program is because in our program, each time a move is made, the entire board is redrawn. This makes it difficult to keep track of who each line on the board belongs to without completely changing the foundational idea of our program. Additionally, when the computer takes a turn in which it completes more than one box, it does so all at once instead of completing the boxes one by one. We did not slow down this process and make the computer wait between each move that it makes because while it always takes a legal turn (set of consecutive moves), the order in which it makes the moves is not necessarily legal. To take our project to the next level, we can take this into consideration.

Furthermore, to improve the performance of our program, we can collect more data on the running of the program against an opponent and conduct more analysis of it. Based on this data, we could try to decipher if there are certain moves or patterns that are evident in all of the cases where the computer loses or ties its opponent. Similarly, we could examine whether all of the computer's wins have some distinguishing factor. If these tests come out conclusive, we could theoretically use the results to make the program truly unbeatable.

Another approach that we could try for our program is incorporating artificial intelligence. This would allow for the possibility of having the computer adapt its strategy and gameplay to the player. Using machine learning strategies, we could show the AI numerous games and have it learn from them. Once it is asked to play against an opponent, it should be able to apply the knowledge that it has gained to play well against the opponent. This could also play a role in decreasing computational time, as the computer would not need to use the minimax algorithm and check every branch of the decision tree for every move.

## VIII. Acknowledgements

We would like to begin by expressing our gratitude to the PGSS Alumni Association and the PGSS Corporate Sponsors for giving us the opportunity to conduct this project. We also want to extend our appreciation to Carnegie Mellon University for providing us with resources and a location to cultivate our skills and knowledge. We would also like to express our utmost thanks to Dr. Barry Luokkala and Ms. Melissa Lessure for taking care of everything behind the scenes. Additionally, we are incredibly grateful to our team project advisor Dr. Neil Simonetti and our teaching assistant, Tripp Hoover for the constant support and advice they provided throughout the duration of the 5 weeks. We are very thankful for the chance to

participate in this thought-provoking project, and we cannot thank the PGSS faculty enough for making this possible.

## IX. References

<sup>1</sup>*CPSC 352 -- artificial intelligence notes: Minimax and alpha beta pruning.* (2011, October 18). Retrieved

August 2, 2023, from

<http://www.cs.trincoll.edu/~ram/cpsc352/notes/minimax.html#:~:text=Alpha%20Pruning%3A%20Search%20can%20be,any%20of%20its%20MIN%20ancestors>

<sup>2</sup>Davis, M. D. and Brams, . Steven J. (2023, June 20). game theory. Encyclopedia Britannica.

<https://www.britannica.com/science/game-theory>

# Paper Soccer: Exploration of Decision Tree Algorithms

Vipin Bhat, Edward Mei, Prakhar Singh, Darren Wang, Eric Xu

## Abstract

Since the inception of artificial intelligence and other advanced algorithmic methods, algorithms have been used to strive toward optimal gameplay in turn-based games. For nearly all popular turn-based games, AI has surpassed the capabilities of human players. The goal of this project was to use different algorithmic approaches such as Minimax, Alpha-Beta Pruning, and the Monte Carlo Tree Search to develop a bot with optimal play in Paper Soccer.

## I. Introduction

### A. Introduction to Paper Soccer

#### 1. Rules

Paper soccer is played on an 8-by-12 grid with two goal extrusions on either end, as shown in Figure 1. At the start of the game, a ball is placed in the middle of the pitch, and players take alternating turns trying to score the ball into the opponent's goal. On each turn, a designated player can move the ball to an adjacent vertex (in one of eight cardinal directions), thus connecting the two vertices by an edge. If the player lands on a vertex that already has at least one prior edge connected to it, the player gets an additional "bounce", meaning their turn continues (see Figure 2 for a turn that incorporates a bounce).

When a player lands on a vertex with no prior connections, their turn ends. As Figure 3 shows, players can also bounce off walls, though it should be noted that "wall-running" is illegal, meaning the player cannot traverse directly up a wall to gain unlimited "bounces". It is also illegal to create a new edge to a vertex on top of an existing edge (that is, connections can't be repeated). A player can win in one of two ways: either by reaching the opponent's goal and scoring or by forcing the opponent to make a move which results in them having no possible legal moves. Scoring a goal is depicted in Figure 4.

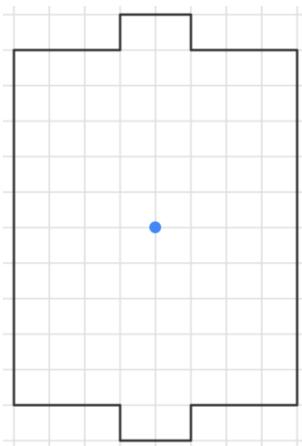
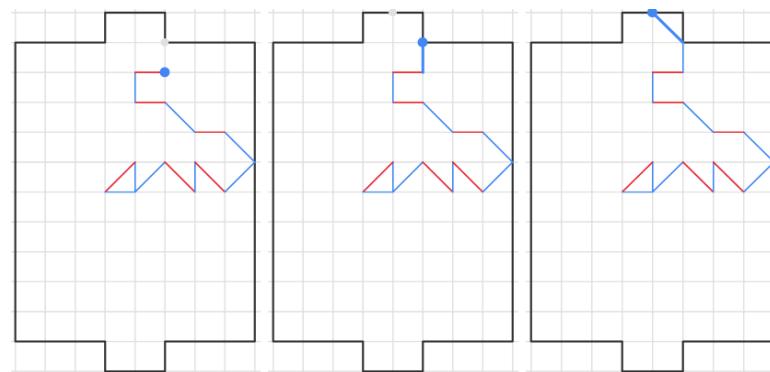
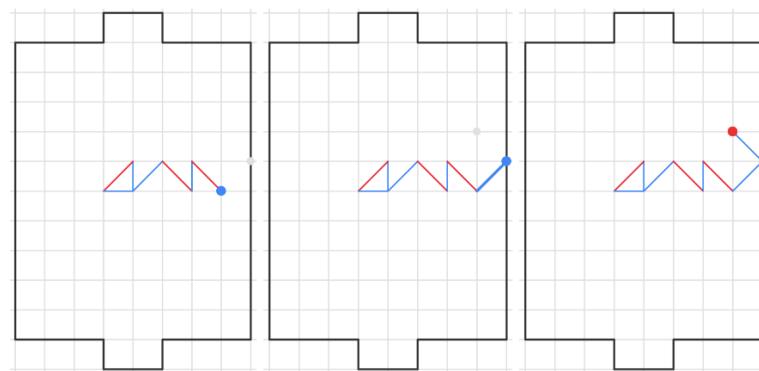
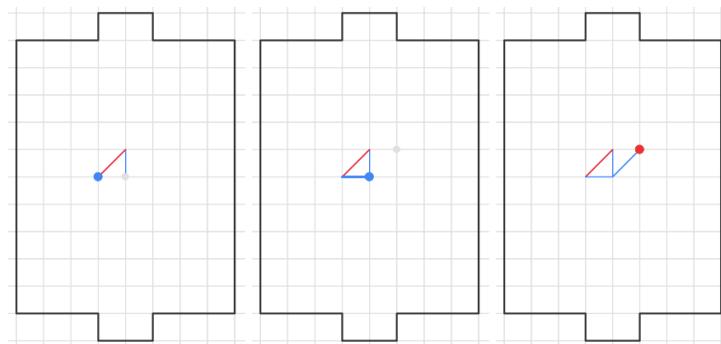


Figure 1: Image of Empty Board

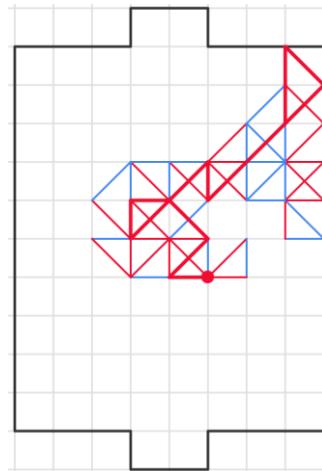
**Figure 2: Vertex Bounce****Figure 3: Wall Bounce****Figure 4: Scoring a Goal**

## 2. Strategies

Paper soccer is a unique game in that it does not adopt a move-by-move format. Instead, players can create combinations of moves so long as the ball (the current point position) touches a vertex that the player can “bounce” off of. The possibilities are endless: turns can range from a simple 2-move kick to a long chain of moves spanning across the board. This distinct game characteristic can lead to the usage of many interesting strategies such as “bouncing”, “boxing-in”, and “goal-blocking”.

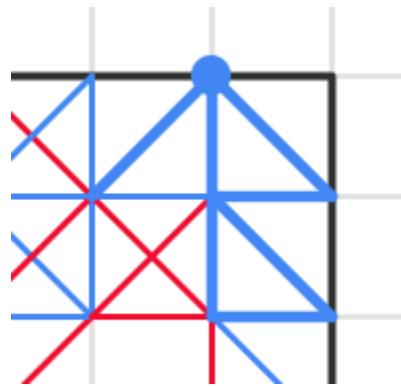
The most prominent strategy in paper soccer is the heavy utilization of “bounces”. As stated in the previous section, bounces allow the player to create long chains of moves that can advance the ball closer to the

opposing player's goal. To pursue this strategy, attacking players often search for and play moves that connect visited vertices in order to push the ball further. Figure 5 depicts a sequence of moves that involves several consecutive bounces. Defending players often attempt to counteract this strategy by playing moves that prevent the opponent from making such bounces, namely by avoiding wall points or previously visited vertices.



**Figure 5: Example of a bouncing chain by the red player**

A more aggressive strategy called “boxing-in” involves trapping the opponent in a way that leaves the opponent with no legal moves left to make. This most often occurs in the corner sections of the board where the attacker “boxes-in” the opponent such that the opponent cannot escape without traversing over an existing connection, leading to a win. One such example is shown in Figure 6.

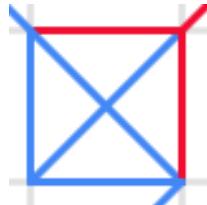


**Figure 6: “Boxed-in” player**

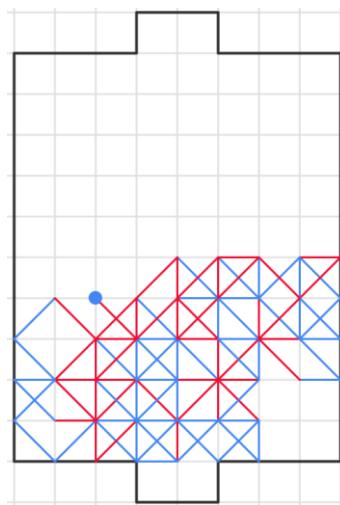
One complicated yet powerful defensive strategy is “goal-blocking”, where a player attempts to block off their goal by forming a series of connections surrounding the goal. Goal-blocking involves the creation of “box-X” structures (refer to Figure 7 for the layout of a “box-X” structure), which prevent the opponent from getting to the opposite side of the “box-X” (since all connections are occupied). If the defending player can create a continuous arrangement of “box-X” structures around their goal, the opponent, unable to repeat connections, is thus rendered unable to enter the goal.

Usually, the defending player allows the attacker to get close to their own goal, but then cleverly utilizes wall “bounces” to create a defensive shield to prevent all future attacks on goal. Figure 8 shows a board

where it is impossible to enter the bottom player's goal. Although goal blocking requires intense positioning to avoid losing the game entirely, it greatly reduces the burden on the player who successfully employs the "goal-blocking strategy".



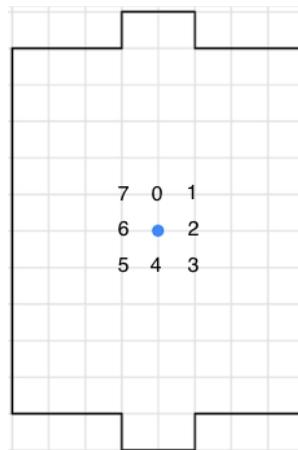
**Figure 7: Box-X structure**



**Figure 8: “Blocked” goal**

### 3. Notation

Each move in paper soccer is notated by a series of numbers, with each number representing a cardinal direction. Figure 9 shows the specific notation. Starting with 0 representing north, numbers 1-7 in ascending order represent the other directions (moving clockwise).



**Figure 9: Paper Soccer Move Notation**

For example, a one-move turn directly north would be notated as “0”. If a player executed a series of bounces moving north, west, and northeast, respectively, then the corresponding notation would be “061”. During a game, players often go back and forth exchanging intricate move combinations, which are separated with a space in-between. For instance, if the first player executed a move-combination in the north-direction followed by the east-direction, and the second player made a responding move-combination in the southwest-direction followed by the south-direction, the proper notation would be “02 54”. Using simple notation to represent move directions allows for more simple implementation, which avoids unnecessary run-time and complexity.

## B. Intro to Algorithms

### 1. History

Algorithms are everywhere in the digital world. They are used in virtually every computer program and can perform tasks as trivial as finding the highest element in a numerical array or tasks as complex as determining a person’s location down to a margin-of-error of only 16 feet using general relativity and advanced mathematics<sup>1</sup>. Despite the seemingly endless complex applications of algorithms, they are merely a series of specific instructions that can be used to solve a problem.

In Babylonian mathematics (dating back to 2500 BC), concrete instructions were used to perform mathematical operations. One of these early algorithms could calculate the square root of numbers by hand and is still used to this day as a common way to calculate roots of numbers without a calculator<sup>2</sup>. Over time, algorithms became more sophisticated and were utilized to find the greatest common divisor of two numbers and decode encrypted messages.

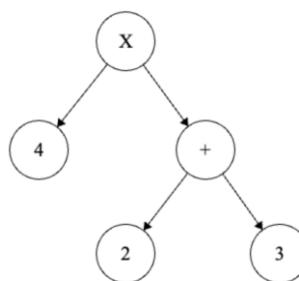
The world’s first published computer algorithm was developed in 1843 by Ada Lovelace to compute Bernoulli numbers on a hypothetical Babbage Machine. Since then, the world of computer algorithms has expanded greatly, and new algorithms continue to be developed that attempt to solve the world’s most complex problems.

### 2. Implementation

To implement an algorithm, a coding language is used to convert abstract instructions into executable actions by a computer or machine. In order to achieve this, a computer uses a system of lexing, parsing, type-checking, linting, and interpreting when reading through a computer program.

When a computer reads source code from a program, it first lexes through this text and creates individual nodes that are then used to represent common operations. These nodes can represent integer values, operators like addition and division, or other special features such as strings or parentheses.

Once these nodes are created, an abstract syntax tree is constructed according to the order they are placed in the source code. As seen in Figure 10, the abstract syntax tree can represent the mathematical statement  $[4 \times (2 + 3)]$  and allows the computer to descend through the tree and perform the operations the way the programmer intends it to do. However, an additional system is required to ensure that the syntax of the source code is valid.



**Figure 10: Abstract syntax tree for the mathematical operation  $[4 * (2 + 3)]$**

This is where type-checking comes into play. Type-checking is a process that investigates the nodes in the abstract syntax tree and ensures that they are valid computational statements. For example, the statement “*Cat* × 2” does not make mathematical sense, so there must be a way to invalidate this statement. Type-checking does this by verifying the types of the nodes that are descended from an operator node and ensuring they are the correct type according to the operator. Linting is another correcting step, where syntax and stylistic errors are adjusted to ensure the interpreter can properly execute the operations the user desires.

Algorithms have become ubiquitous throughout both industry and research. They are commonly used as optimization techniques, evaluating situations much more efficiently than humans could. For example, the powerful chess engine Stockfish can evaluate up to 70 million positions each second<sup>3</sup>, well beyond what even top grandmasters can accomplish.

The new face of computer algorithms is deep learning, a subset of artificial intelligence (AI) that attempts to emulate the intricate connections of the human brain. For example, the neural network-based chess engine AlphaZero, developed by DeepMind, uses reinforcement learning in conjunction with Monte Carlo Tree Search. AlphaZero trains by playing against itself, gradually improving each time by using prior game data. The use of a neural network allows AlphaZero to hone in on promising positions, and thus is able to achieve comparable performance to Stockfish while only evaluating 80,000 positions, nearly a 1000 times fewer than Stockfish. These techniques also allowed AlphaGo, a specialized form of AlphaZero, to defeat a professional world champion-class player in Go, a game with around  $10^{170}$  legal positions.

AI is also commonly used for scientific research. AI tools can greatly help researchers in streamlining and/or automating repetitive tasks. AI can also play essential roles in novel discoveries. AlphaFold2, an AI program that predicts protein structures from amino acid sequences, is already playing a substantial role in drug development.

## II. Setting up the Game

### A. Development Environment

#### 1. Language

The object-oriented programming language Java was used for its platform-independent programming style, which ensured the program could run properly regardless of the environment. The design of Java also emphasized simplicity and robustness which made it highly suitable for the project. In addition to the features Java offers, Java is a well-known programming language that most of the team had prior

experience with, greatly reducing the amount of time dedicated to instruction and avoiding the need to learn a new programming language.

## 2. Tools

To code the game and bot, Replit and Eclipse were used as integrated development environments (IDE). Replit is a well-known online IDE that allows for rapid sharing of programs and real-time collaboration on code. This allowed the entire team to work on one singular project without having to locally host the program on each of their machines or use additional software like GitHub to enable version control. However, Replit came with major drawbacks, especially with run-time performance. Since Replit uses virtual resources rather than local hardware, it had significantly fewer capabilities and would often cause the game to lag or crash, preventing effective testing.

The solution was to use Eclipse IDE simultaneously with Replit. Eclipse was used to work locally, with any updated code being transferred over to the shared Replit. This allowed the team to work with greater computer resources and additional tools that made the programming process much more efficient with less lag and fewer crashes.

## B. Program Structure

### 1. Object-Oriented Programming (OOP)

Object-oriented programming is a common style of programming in languages such as Java, Python, Ruby, and many others. In these types of programming languages. This programming-language structure is widely used due to its advantageous properties, those being inheritance, polymorphism, encapsulation, and modularity.

In Java, every variable is an instance of a class. Classes serve as a blueprint to define the structure and behavior of these objects. Each object has its own field variables and methods, each providing crucial data about said object or performing operations on those objects. Field variables can provide pertinent data detailing class properties or fixed values, while methods can perform operations from inputs or field variables.

Inheritance allows sub-classes to inherit properties and methods from more-general classes (superclasses) without needing to rewrite the same fields and methods in the sub-class. For example, an Eagle object class can inherit methods and field variables from a more-general Bird object class, meaning that the Eagle object class does not need to be rewritten and can instead simply acquire these class traits from its Bird superclass. See Figures 11 and 12 for the specific syntax behind superclasses and subclasses.

```
public class Bot {  
    public Bot() {  
    }  
}
```

Figure 11: Bot Superclass

```
public class ControlAlgorithmBot extends Bot{  
    public ControlAlgorithmBot() {  
    }  
}
```

Figure 12: ControlAlgorithmBot Subclass

Polymorphism enables the objects of different classes to be treated as instances of a common superclass, promoting flexibility in the code. This is similar to inheritance, but is slightly different in that polymorphism allows subclasses to exhibit slight changes in a common method that is inherited from their superclass.

Encapsulation ensures that data within an object is hidden and can only be accessed through predefined methods (often known as getter and setter methods) to ensure data security and integrity. This is done by making fields and methods private or public. If a field or method is private, it renders it inaccessible by exterior classes and is only accessible within the class. For example, the variable “isPlayerTurn” as seen in Figure 13 is a private variable, meaning that value cannot be directly accessed in classes other than its host class. This is often done to ensure that critical data is not interfered with by the user and provides security to the program.

However, if a programmer wants to make a field or method available for the user to utilize, that said field or method can be assigned “public” so that it can be accessed in other classes. In Figure 14, the method getPlayerTurn() is public, meaning that the method can be directly called from other classes.

```
private boolean isPlayerTurn;
```

Figure 13: Private field variable

```
public boolean getPlayerTurn() {  
    return isPlayerTurn;  
}
```

Figure 14: Public method

The modularity of the Java language also allows the project to be divided into various classes that can then be put together in one main class to create a complex yet easy-to-comprehend program. Modularity was a highly-desired feature in the programming language chosen as it granted the ability to split the program into various different classes such as Board, Move, Animation, and Bot which could be worked on independently by individual team members.

## 2. Animations/GUI/Java Swing

Animations in GUI with Java Swing make the program more interactive. Java Swing is a powerful Java framework that allows developers to incorporate smooth and engaging animations into their applications. Java Swing's versatility and ease of use enable developers to create dynamic and appealing interfaces.

All Java Swing components such as buttons, labels, text fields are part of the “JComponent” class. This allows for it to be extended to customize components for specific requirements. It is also lightweight, meaning that it runs smoothly even on lower-end hardware and has a consistent feel across all platforms such as Windows or MacOS. The layout managers in Java Swing help in arranging components within containers such as frames, panels, and dialogs. They dynamically adjust the layout based on the container's size. Swing applications are event-driven where actions with components trigger events. Listeners and handlers are used to respond to these events such as mouse clicks or movement.

Code-wise, animations are implemented with a copy of the game board that has continuous rather than discrete values. These continuous values have inertia and are “rubber-banded” to their discrete counterparts, creating smooth motion that helps to convey relationships between UI elements.

## 3. Data Collection

A large database of previous games played was available at playok.com, a website that hosts online paper soccer games. Hundreds of thousands of games were available for download, providing a sizable amount of training data for the team's AI algorithms. Game data was downloaded in the form of .txt files, which were then converted into readable numerical data in Excel spreadsheets. This was accomplished using Python and supplementary libraries. Processing was done to remove extraneous information and unwanted data. The data also recorded moves using the notation stated in section I.A.3, which allowed the programs to conveniently parse through the game data from playok.com without inconveniently needing to convert between notation systems.

# III. Algorithms

## A. Control Algorithm

### 1. Process

The control algorithm simply picks a random move from the list of legal moves. This list of moves is procured by a method called validMoves() that takes in x and y coordinates (the position of the ball) and checks each adjacent vertex to see if a move there is legal. If the random move happens to give the algorithm an extra bounce, the algorithm is able to recognize this and continue making moves until its turn is over. However, the control algorithm has no preference for a particular type of move and cannot purposefully make moves with bounces. It also lacks an inherent sense of direction, so it sometimes scores on itself. The control algorithm was used as a baseline that more advanced algorithms could be compared to.

### 2. Performance and Simulations

Although the control algorithm is a reliable way to simulate random moves, unfortunately it is not a strategic algorithm and loses quite frequently when faced against more formidable opponents such as GreedyAlgorithmBot, MinimaxBot, and MCTSBot (Monte Carlo Tree Search Bot). However, creating a

control algorithm also allowed other interesting properties of the game to be investigated, namely the possibility of a first-turn advantage in paper soccer.

By using a class called BatchBoard(), large numbers of simulations of bot-versus-bot games can be created between various bot types. To test the possibility that the first player has a higher win probability than the second player, an identical intelligence simulation was conducted between two control algorithm bots for 1000 trials.

In Table 1, data regarding the win-rates wins was collected.

**Table 1: Data from Control Algorithm Bot vs. Control Algorithm Bot**

Number of Simulated Games	Number of Control Algorithm Bot #1 Wins (Moves First)	Number of Control Algorithm Bot #2 Wins (Moves Second)
1000	518 (51.8%)	482 (48.2%)

After collecting data from the 1000 trials of games with Control Algorithm Bot versus Control Algorithm Bot, a statistical 1-sample Z-test of proportions was used to test whether or not there is a statistically significant first-turn advantage in paper soccer. The results of this test are shown in Table 2.

**Table 2: 1-Sample Z-test data from above Table 1 simulation**

P-Value for ( $\neq P$ )	P-Value for ( $> P$ )	Significance Level
0.25495	0.12747	0.05

Because both p-values are greater than the significance level of 0.05, the null hypothesis that there is not a first-move advantage fails to be rejected. This means that despite these seemingly strong results, there is not enough convincing statistical evidence that the player beginning the game has a different chance at winning the game than the second player.

The results from this Z-test are important as they can help attribute the other simulation results solely to bot strength and not to the order the bots played in. Z-tests were not conducted on later algorithms as the calculated p-values were always near zero, meaning the results can be attributed to the algorithm strength rather than chance.

## B. Greedy Algorithm

### 1. Process

The greedy algorithm solely focuses on proximity to goal, without worrying about getting bounces or other strategies. This algorithm always picks the move that goes up toward the opponent's goal. If it can't find such a move, i.e. all upward moves are illegal, the algorithm resorts back to a random process as described in the aforementioned control algorithm.

The greedy algorithm also uses a simple distance formula, where the absolute value of the difference between the horizontal center of the board and the final x-coordinate of each move is calculated. This

formula is utilized by the bot in order to try to stay close to the center of the board horizontally and minimize this distance to the opposing goal. That is, upward moves that stay closer to the center are preferentially selected over upward moves that move away from the center.

## 2. Performance and Simulations

Once implemented, the greedy algorithm performed considerably better than the control algorithm, but still faced challenges against more powerful bots such as MinimaxBot. By performing an analysis in matches between ControlAlgorithmBot and GreedyAlgorithmBot, the greedy algorithm's strength relative to the control algorithm can be determined. Additionally, the starting order of the bots was also alternated to ensure the win-rate results were determined by algorithm strength only.

**Table 3: Data from Control Algorithm Bot (First) vs. Greedy Algorithm Bot (Second)**

Number of Simulated Games	Number of Control Algorithm Bot Wins (Moves First)	Number of Greedy Algorithm Bot Wins (Moves Second)
1000	29 (2.9%)	971 (97.1%)

**Table 4: Data from Greedy Algorithm Bot (First) vs. Control Algorithm Bot (Second)**

Number of Simulated Games	Number of Greedy Algorithm Bot Wins (Moves First)	Number of Control Algorithm Bot Wins (Moves Second)
1000	985 (98.5%)	15 (1.5%)

As shown in Tables 3 and 4, the greedy algorithm shows considerable strength over the control algorithm and can reliably win against it in most cases. However, there were a few losses that arose in these matches as seen in the data above. During testing it became clear that although the greedy algorithm could find the closest path to the opposing goal, it could not account for cases where it would get entrapped while bouncing towards the goal. This resulted in a small number of losses, but the greedy algorithm still remains the more competitive algorithm when compared against the control algorithm.

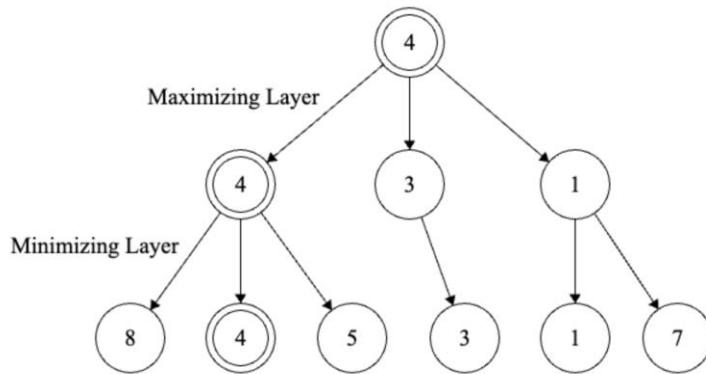
## C. Minimax Algorithm and Board Scoring

### 1. Process

The minimax algorithm relies heavily on recursive processes and a method that evaluates board positions<sup>4</sup>. This “board scorer” returns a value between 0 and 1, with 0 representing a board that is winning for the human player, and 1 representing a board position that is winning for the algorithm. The manually-created method treats the score as a function of the y-position of the ball and the number of bounces available. Bounces that were available by moving forward were rewarded more heavily than bounces that were available by moving on the same level or by moving backward. The more bounces a turn contained also increased the score of the board towards the algorithm. In addition, the bot also searches for and punishes moves that would result in entrapment in its score calculation.

Another characteristic of minimax is depth, which communicates to the algorithm how far into the future to search. Theoretically, greater depth should yield higher playing performance because it is seeing more moves ahead. It should be noted that minimax's performance (in terms of run-time) decreases significantly with increasing depth. Because of the recursive nature of minimax, an increase in depth results in an exponential increase in the time/number of operations to be completed. Therefore, it is unrealistic to rely solely on a large depth to achieve high-level and time-efficient play. A proper balance between bot intelligence and run-time efficiency is required.

Given a board position, minimax looks at each board position that would result from each possible move. This continues until the maximum depth is reached, with minimax being recursively called on each board position. Once the `minimax()` method is called, a tree of possible board positions is created as seen in Figure 15. The nodes at the very bottom of the tree are then evaluated by the board scoring method. At each level of the tree, a minimizer or maximizer operation is conducted: an optimal human player will try to minimize the score, while the algorithm will try to maximize their score. Therefore, it is possible to trace up the tree from each board at the bottom and determine what outcome each move will have, assuming optimal play. The algorithm takes in this information and chooses the move that will result in the highest board score for the algorithm.

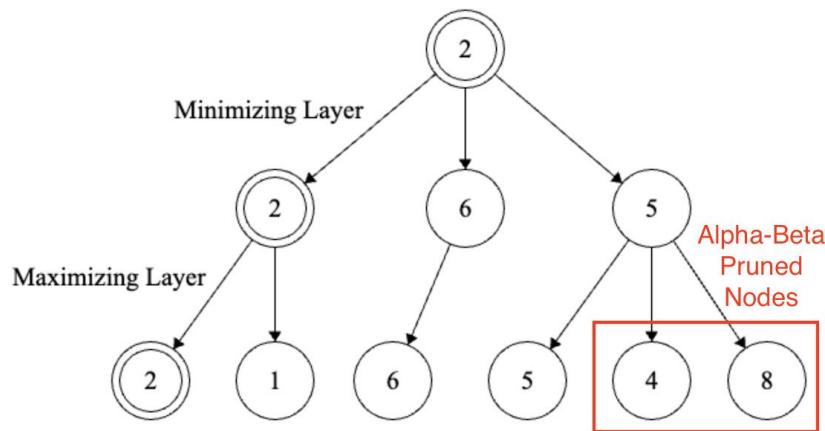


**Figure 15: Minimax Search Tree with a Maximizing Layer Followed by Minimizing Layer**

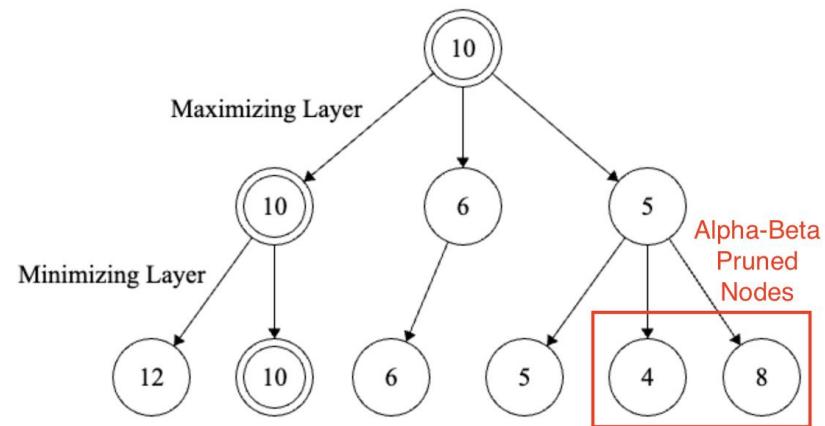
## 2. Alpha-Beta Pruning and Trees

To combat run-time inefficiency, alpha-beta pruning is used to reduce the number of future boards the minimax bot needs to check. Consider the case where the minimax algorithm is on a minimizing layer (player's turn) and can currently guarantee a minimum value *beta* based on the child nodes it has already looked at. When minimax is called on the subsequent child node, the value that is returned is the maximum of the children of that child node.

However, the original minimizer is only interested in maximum values that are less than *beta*. Therefore, in the minimax call on the child node, the function can immediately exit any time the currently maintained maximum value becomes greater than or equal to *beta*, culling out nodes before they are even processed. These nodes are known as alpha-beta pruned nodes (as seen in Figures 16 and 17). Similar logic can be applied when the minimax algorithm is on a maximizing layer and can currently guarantee a maximum value *alpha*. By maintaining the values *alpha* and *beta* throughout the minimax algorithm, the number of node visits the algorithm performs can be cut down significantly.



**Figure 16: Minimax Search Tree with Minimizing Layer Followed by Maximizing Layer and Alpha-Beta Pruning**



**Figure 17: Minimax Search Tree with Maximizing Layer Followed by Minimizing Layer and Alpha-Beta Pruning**

Overall, alpha-beta pruning allows the minimax algorithm to avoid descending through unnecessary nodes which will always return a score that is not relevant to the most-optimal move set. While the algorithm does not change the time complexity of minimax, it still reduces the run-time of minimax by a significant factor.

### 3. Performance and Simulations

Minimax algorithms are incredibly powerful algorithms when it comes to artificial intelligence-based game bots, but sometimes that power comes at a cost. In the case of minimax algorithms, run-time efficiency is a major weakness. This is because of its recursive nature, looking  $N$  turns into the future (if recursion depth =  $N$ ), scoring the outcome of every possible future board position  $N$  turns in the future, and then choosing the most-optimal board moves to play to achieve the highest possible score from those.

One observation was that the higher depths of the minimax algorithm would result in the program running slower. This is likely due to the enormous number of future board positions possible with each pair of turns and expands exponentially, causing run-time to also increase exponentially. Despite the relatively longer

run-times, the minimax algorithm consistently won against the control and greedy algorithms, even at depth = 1 (the lowest depth possible). While creating the minimax algorithm bot, a separate parameter called depth was implemented, which indicates the amount of turns the bot looks into the future (e.g. MinimaxBot(2) would have a depth of 2).

Tables 5 and 6 detail the specific 1000-trial matchups of MinimaxBot against ControlAlgorithmBot and GreedyAlgorithmBot, with the minimax algorithm having a depth of 2.

**Table 5: Data from Minimax(1) Algorithm Bot [First] vs. Control Algorithm Bot [Second]**

Number of Simulated Games	Number of Minimax(1) Algorithm Bot Wins (Moves First)	Number of Control Algorithm Bot Wins (Moves Second)
1000	1000 (100.0%)	0 (0.0%)

**Table 6: Data from Minimax(1) Algorithm Bot [First] vs. Greedy Algorithm Bot [Second]**

Number of Simulated Games	Number of Minimax(1) Algorithm Bot Wins (Moves First)	Number of Greedy Algorithm Bot Wins (Moves Second)
1000	1000 (100.0%)	0 (0.0%)

These simulations demonstrated that the minimax algorithm could definitively win against the control algorithm and greedy algorithms, regardless of which bot started the first move. This establishes that the minimax algorithm is a more competitive algorithm than both the control and greedy algorithms, primarily due to the comprehensive searches the minimax algorithm conducts.

## D. Monte Carlo Search Trees and k-Nearest Neighbors

### 1. Process

The Monte Carlo Search Trees (MCTS) with k-Nearest Neighbors (kNN) is a powerful combination for implementing an intelligent agent in paper soccer. The process involves combining the exploration and exploitation capabilities of MCTS with the pattern recognition of kNN to create an adaptive yet competitive player.

MCTS is a simulation-based search algorithm that repeatedly performs the following steps: selection, expansion, simulation, and back propagation<sup>5</sup>. This iteratively builds up a game tree. During the selection step, the algorithm chooses a path to descend down the tree. The path is chosen by the Upper Confidence Bound formula, which attempts to balance exploitation (visiting of nodes with high win rates) and exploration (visiting of nodes with relatively low visiting rates). That is, the algorithm tries to select both promising nodes on the tree and nodes that haven't been visited often. At the end of the path, a new node is created. The exact means by which MCTS chooses to descend down the tree and expand to new nodes is managed by a formula for evaluating nodes known as the Upper Confidence Bound (UCB) formula, which MCTS attempts to maximize when descending/expanding to new nodes. The formula is given in Equation 1.

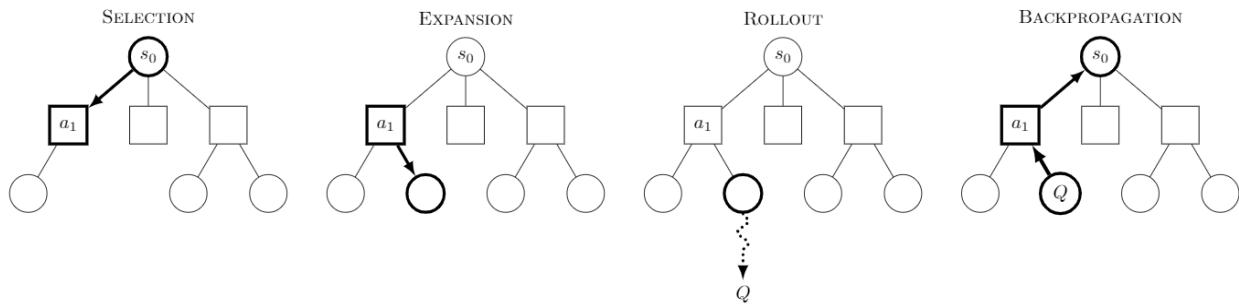
$$UCB(node_i) = \bar{x}_i + c\sqrt{\frac{\log N}{n_i}}$$

$\bar{x}_i$ : mean node value;  $n_i$ : #visits of node  $i$ ;  $N$  #visits parent;

**Equation 1: The Upper Confidence Bound Formula, also known as UCB.<sup>5</sup>**

In the formula, the first term represents the expected number of wins for a node (approximated by the rollout phase, as will be mentioned later), while the second term becomes larger when a node is relatively unvisited. Thus, when MCTS attempts to maximize UCB, it takes into consideration which nodes have high win rates and which nodes are rarely visited, ultimately balancing exploitation and exploration. As an added nuance, MCTS must use the loss rate instead of the win rate in the UCB formula when it is on every second layer in the game tree, reflecting the fact that every other layer in the game tree represents the opponent player who wants to maximize losses. Additionally, the UCB formula can be tuned using the C factor, which determines how much weight to give to the exploration factor as compared to the exploitation factor.

After the expansion phase, one or more random simulations are played out from the node that was created, where the result (wins or losses) is saved. This is known as the rollout phase. Statistically speaking, a node that represents a winning position is more likely to result in a win than a loss, and this is expressed in the results of the random simulation. Lastly, the results of the random simulation are propagated back up the tree. This way, all of the nodes upstream of our newly expanded node know whether the new node is promising or not. By iteratively applying the four phases of selection, expansion, rollout, and backpropagation, the MCTS algorithm builds up a game tree that can later be used to generate a sequence of moves that constitutes a turn. See Figure 18 for an overview of this process.



**Figure 18: Four Phases of the Monte Carlo Tree Search Algorithm<sup>6</sup>**

The key strength of traditional MCTS is that it only requires the rules of the game that it is applied to. The algorithm figures out winning and losing positions based on the rollout mechanic, which as mentioned earlier uses nothing but random choice. As such, it has very wide-reaching applicability and is often what programmers turn to when approaching an unfamiliar game.

However, the randomness aspect of MCTS can also be a downside, especially when algorithms exist to make more informed choices given a board state. For example, a neural network called a policy network can be introduced to take an existing board state and output a suggestion for possible moves to be made (see section V.B.2)<sup>7</sup>. The introduction of informed choice can provide a powerful addition to MCTS that can make it competitive with expert humans.

An alternative solution towards suggesting moves given a board state would be to use kNN, or k-Nearest Neighbors. kNN requires an existing dataset of board states labeled with suggested moves. In the actual implementation, the board states are converted into vectors (points in high-dimensional space) before being placed in the dataset. When given a board state converted into a vector, kNN outputs vectors in the dataset that are proximally close to the query vector. The corresponding moves for the vectors are then inputted into a custom-made consensus algorithm which outputs a suggestion for a move. To deal with large datasets, kNN uses a heuristic algorithm known as LSH (Locality Sensitive Hashing), which hashes vectors in a way such that close vectors have similar hashes<sup>8</sup>. The implementation of LSH was taken from the Java library known as SMILE (Statistical Machine Learning & Learning Engine).

## 2. Implementation with kNN and Trees

To create board representations to coincide with kNN, each board was turned into a vector of length 939. This vector contained true/false values for every possible combination for each point on the board, with 1 and 0 representing true and false. Additionally, the ball's x and y coordinates are added to the vector. Because this is much more important in comparing board similarity than a simple connection, they are weighted much more heavily. Lastly, another weighted component that described player turn was added. The full form of this vector can be found in Figure 19. Each of these board representations was created by processing data taken from playok.com. Again, Python was used to interact with and format the data. In total, this amounted to ~4000 games and ~400,000 board vectors. Each board representation was also associated with the next move played in the actual game. Therefore, using these associations, it is possible to look at boards to predict the best next move. To increase the amount of data available to analyze, board representations were taken after every individual move, not after every overall move. In the case of the move 184, board vectors were created after move 1, after move 8, and after move 4, rather than creating a single vector after move 184.

To find the k nearest neighbors, board vector representations' similarity were compared using euclidean distance. After finding the nearest neighboring boards, the move associated with each board representation was retrieved. These boards then "voted" on what next move to make, with closer boards having more voting power than less similar boards. Because of inefficient run times, it was unrealistic to use all 400,000 vectors. Instead, only 10,000 were used in the practical implementation due to computer performance restrictions.

**Figure 19:** Example of a 939-term vector describing a board state

## IV. Results

### A. Comparative Analysis

## 1. Trial Outcomes: Bots versus Bots

Bot-versus-bot trials were critical in determining which algorithm was the most competitive regardless of run-time. 1000 simulated games were conducted, and the win-rate results were obtained. These results are summarized in Table 7. Identical-intelligence bots were not faced against each other in this comparative analysis.

**Table 7: Bot-versus-bot matchup graph with 1000 trials [First Player Wins : Second Player Wins]**

		<b>Second Move Bot</b>		
		Bot Type	Control Algorithm Bot	Greedy Algorithm Bot
<b>First Move Bot</b>	Control Algorithm Bot	X	15 : 985	0 : 1000
	Greedy Algorithm Bot	971 : 29	X	0 : 1000
	Minimax Algorithm Bot (Strength = 2)	1000 : 0	1000 : 0	X

## 2. Recursion Depth and Effect on Win Percentage

Another possible factor that could affect win percentage is the depth of the recursive minimax algorithm. The depth (also called strength) determines how many minimizing and maximizing layers the minimax game tree contains. This depth also determines how many moves in the future the minimax algorithm bot investigates. To test the effects of the strength of the minimax algorithm bot, matchups of minimax at varying depths were conducted.

Based on the data above, it was concluded that generally, higher minimax depth/strength corresponded to a higher win rate. For example, Minimax(2) always defeated Minimax(1) and Minimax(3) always defeated Minimax (1). Originally, it was believed that this trend existed because the additional depth would allow the minimax algorithm to look further ahead into future turns and make more accurate turns that would benefit its chances of winning.

However, one notable exception was the matchup between Minimax(2) algorithm bot and Minimax(3) algorithm bot, where the Minimax(2) algorithm defeated Minimax(3) every time out of 1000 trials. While this finding was surprising at first, it was observed that minimax algorithm bot vs. minimax algorithm bot games were deterministic. This meant that no matter how many simulated games were played, the same outcomes would always emerge from every single one of the 1000 games since each minimax would make the same exact move sequence. Because minimax chooses the mathematically best outcome, it is expected that when these two minimax algorithms compete against each other, they choose the same set of moves.

In the case of the Minimax(2) versus Minimax(3), the discrepancy in the depth vs. win-rate trend could be attributed to the incomplete representation of a board's evaluation by the board scoring method. The board scoring minimax algorithm implemented only bases its evaluation on the distance of the ball to the opposing goal and the number of bounces that the ball could possibly make. However, there may be other factors that can influence the favorability of a board in paper soccer such as the number of nearby connections and the extent to which the user's goal is blocked.

### 3. MCTS Results

Trial games with Monte Carlo Tree Search Bots (MCTSbots) were also conducted. Comparative analysis was done with two main categories of MCTS: traditional MCTS (MCTS-Random) and MCTS with kNN (MCTS-kNN). In order to optimize MCTS-Random, the algorithm played itself but with different values of the constant C (see equation 1) in the UCB formula. The results are described in Table 8, with each matchup given 100 trials:

**Table 8: Matchup graph with different C-values for MCTS**

		Second Move Bot			
		Value of C used for MCTS	10	1	0.1
First Move Bot	10	X	34 : 66	89 : 11	20 : 80
	1	66 : 34	X	14 : 86	32 : 68
	0.1	94 : 6	91 : 9	X	72 : 28
	0.01	76 : 24	69 : 31	38 : 62	X

The results imply that the best value for C is approximately C = 0.1. This value was then used for all future applications of MCTS-Random and MCTS-kNN. MCTS-Random was then played against MCTS-kNN in batches of 20 trials (due to the long amount of computation time that MCTS-kNN needs) and the results were aggregated in Table 9.

**Table 9: Matchup Graph between MCTS-Random and MCTS-kNN**

		Second Move Bot	
		MCTS-Random	MCTS-kNN
		Bot Type	
First Move Bot	MCTS-Random	X	7 : 13
	MCTS-kNN	12 : 8	X

These results show that MCTS-kNN has a slight win-rate advantage over MCTS-Random. However, MCTS-kNN is much less time-efficient, taking significantly more time to compute a turn than MCTS-Random. From an overall analysis, the two algorithms likely lie on a similar level of competitiveness, which was not completely expected. However, the MCTS-kNN algorithm has the potential to become significantly more competitive given enough optimization and more training data.

The MCTS-Random was also played against the greedy algorithm and Minimax(2). In each matchup, half of the trials had MCTS-Random play first and half of the trials had MCTS-Random play second to eliminate bias. As MCTS-kNN takes a significant amount of time to compute moves, no trials were conducted for MCTS-kNN. Rather, the MCTS-Random algorithm was chosen to be a representative algorithm for MCTS, since the two versions of MCTS have similar performance. The MCTS algorithm's win rates are summarized in Table 10.

**Table 10: Win rate of MCTS (Random) against different bots**

	Greedy Algorithm	Minimax(2)
MCTS-Random Win Rate	<b>185 / 200</b>	<b>1 / 50</b>

The results imply that MCTS can defeat the greedy algorithm pretty consistently, but is virtually completely defeated by Minimax(2), suggesting that modifications may be needed to bring MCTS to the level of minimax.

## V. Future Research

### A. Reinforcement Learning

#### 1. Purpose

Reinforcement learning is a powerful tool in machine learning that focuses on training intelligent models without the use of big data but rather through trial and error. Unlike other machine learning models, reinforcement relies on interactions with an environment to learn optimal actions. The process of learning mirrors how humans and animals learn from experience.

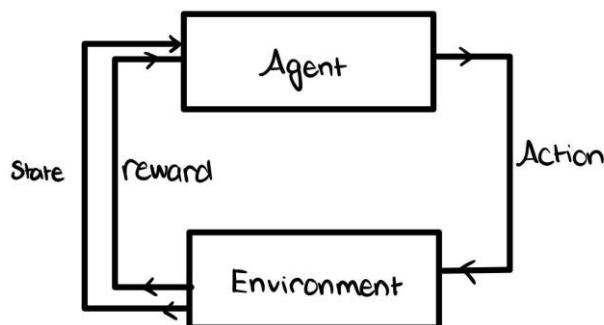
In reinforcement learning, an agent receives feedback from its environment in the form of rewards or penalties for its actions. The objective of the agent is to maximize its rewards over time by finding the best actions. This form of learning is widely used for playing games as it simulates how a human would slowly get better by putting in more and more time into the activity. Games like Chess, and Go, as well as simple Atari games such as Pong and Space Invaders have become a baseline for training reinforcement models.

Reinforcement learning leverages concepts from the Markov Decision Process and dynamic programming to efficiently calculate the trade-offs between exploration and exploitation. Deep reinforcement learning combines traditional reinforcement learning with deep diurnal networks to solve complex problems in high dimensions. Some examples of such networks are Deep Q-Networks and Proximal Policy Optimization.

## 2. Methodology

Implementing reinforcement learning into Paper Soccer involves creating an intelligent agent that is capable of understanding the game and interacting with the environment. It needs to be able to receive feedback based on its actions. Due to time constraints, the team was unable to implement this technique as it takes multiple days of training for a variable model. However, below is the methodology for implementing this technique in the game of Paper Soccer.

Firstly, the environment has to be defined including the game grid, the state of the ball, and the boundaries. Each cell must be marked with valid paths and occupied positions. Next, the State Space has to be defined for a reinforcement learning agent to understand the environment. A possible state would be the current ball position along with the moves of the other player. In addition, an Action Space must be defined which is the set of possible actions the reinforcement learning agent can perform. In this case, this would be the valid paths on the grid and “flicking” the ball in those directions. Furthermore, picking a reinforcement learning algorithm such as Q-Learning would be required for ease of implementation. Finally, a reward would have to be designed that would encourage the learner to score goals. An example of this would be moving towards/scoring a goal would lead to positive reinforcement and getting scored on by another player would lead to negative reinforcement. Figure 20 highlights the key points of the reinforcement learning process.



**Figure 20: Reinforcement Learning Process**

After the setup of the learning agent, the training period would start in which the learner would repeatedly play games against the other AI's such as the Minimax or Greedy. The training can be a time taking process however it is rewarding as usually more training leads to better performing models. Lastly, based on the progress and performance of the model, there may need to be changes implemented into the algorithm to retrain the model.

## B. Artificial Intelligence-Based Board Scoring Using Value and Policy Networks

### 1. Purpose

Currently, the board scoring method involves a weighted measure of the ball's distance to the opposing goal and the number of bounces the ball can make in a particular position. This scoring criterion was decided on after manual analysis of many games and strategies from highly-ranked players (on playok.com), where long series of bounces appeared to be generally positive moves for a player to make.

However, there still exist some inaccuracies within the board scoring method. For instance, the method cannot give board scores based on the presence of a blocked goal (which would highly favor one player over another) and cannot determine board scores based on the entire path that the ball has currently taken. This data is much more complex as such is much harder to analyze for patterns.

## 2. Methodology

To implement an artificial intelligence-based board scoring method, value networks could be employed to produce a more informed score of a position. These neural networks can be trained on existing data of positions pulled from websites like playok.com. However, this method only allows the value network to become as good as the training data, which would be likely based on human evaluations. One possible method to create “superhuman” value networks is through reinforcement learning, introducing a second network called a policy network that predicts moves based on a current board position. By having these two networks improve by learning from each other, a resulting superior value network (and policy network) can be created. In fact, DeepMind’s AlphaGo system used value and policy networks along with the Monte Carlo Search Tree algorithm in order to surpass professional human Go players<sup>9</sup>.

# VI. Conclusion

## A. Summary

### 1. Summary of Experimentation

In conclusion, the Minimax algorithm performed best in the one-on-one tests. Specifically, Minimax with a depth of 2 was able to handily defeat all other bots, including other iterations of Minimax. MCTS-Random and MCTS-kNN finished at roughly the same performance level when taking into account both move accuracy and runtime efficiency. Specifically, MCTS-kNN made more accurate moves but also took a considerably long time to find a move. The MCTS algorithms lost a few games against the greedy algorithm, while Minimax was able to win every time (refer to Figure 21 for the full ranking of algorithms). The team hypothesizes that with a few weight adjustments, greater volumes of training data, or adjustments to the UCB formula, MCTS-kNN could be able to reach greater performance. To enable this, a way to increase runtime efficiency would also be required to offset the inflated amount of data.

1. Minimax

2. MCTS-kNN

3. MCTS-Random

4. Greedy

5. Control

Figure 21: Board Performance Ranking

## B. Acknowledgements

### 1. People

This project could not have been possible without the generous support of Professor Neil Simonetti and teaching assistant Krystof Purtell. Without their constant encouragement and wisdom, the journey we chose

to embark on would have never been as successful and enjoyable. Additionally, the advice from Professor John Urbanic was invaluable to the artificial intelligence portion of our project.

Finally, many thanks to Cameron Blackwood-Short and Dr. Barry Luokkala for all their support throughout the Pennsylvania Governor's School for the Sciences program. We truly appreciate their tireless dedication to supporting high school students in the sciences and continuing this amazing program. We would also like to thank the PGSS Alumni Association for all of their unwavering support of the PGSS Program. All of these amazing people and organizations played a crucial role in the success of this project. Without them, this project wouldn't have been close to fruition.

## **2. Resources**

In addition to all the help the team received from mentors, this project could not have been possible without the generosity of PlayOK, a free website that hosts and records live paper soccer games and tournaments with players around the world. This project also would not have been possible without the ample computer resources and workspaces provided to us by Baker Hall, Hunt Library, Posner Hall, and the Jared L. Cohon University Center, all at Carnegie Mellon University. Each of these locations provided generous resources for our project's growth and allowed us to collaborate meaningfully toward our project's goal. It is for these reasons that we are truly grateful for the opportunities and resources provided to us by PlayOK and Carnegie Mellon University.

## **Appendix: Code**

The GitHub repository storing all of the code used in the project can be found here:  
<https://github.com/FriedRice53/PaperSoccerGameBot.git>

## References

<sup>1</sup> van Diggelen, Frank, Enge, Per, "The World's first GPS MOOC and Worldwide Laboratory using Smartphones," Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015), Tampa, Florida, September 2015, pp. 361-369.

<sup>2</sup> Fowler, David, and Robson, Eleanor. "Square Root Approximations in Old Babylonian Mathematics: YBC 7289 in Context." Historia Mathematica, vol. 25, no. 4, 1998, pp. 366-378, <https://doi.org/10.1006/hmat.1998.2209>. Accessed 4 Aug. 2023.

<sup>3</sup> Silver, David, et al. "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm." arXiv.Org, 5 Dec. 2017, arxiv.org/abs/1712.01815.

<sup>4</sup> Rajiv Bakulesh Shah, "minimax", in Dictionary of Algorithms and Data Structures [online], Paul E. Black, ed. 10 January 2007. (accessed 4 Aug. 2023) Available from: <https://www.nist.gov/dads/HTML/minimax.html>

<sup>5</sup> Wang, Benjamin. "Monte Carlo Tree Search: An Introduction." Medium, 11 Jan. 2021, [towardsdatascience.com/monte-carlo-tree-search-an-introduction-503d8c04e168](https://towardsdatascience.com/monte-carlo-tree-search-an-introduction-503d8c04e168).

<sup>6</sup> Rmoss92. *MCTS-Diagram*, <https://commons.wikimedia.org/wiki/File:MCTS-diagram.svg>.

<sup>7</sup> "Monte Carlo Tree Search MIT Courseware." YouTube, YouTube, 26 Oct. 2018, <https://www.youtube.com/watch?v=xmImNoDc9Z4>. Accessed 4 Aug. 2023.

<sup>8</sup> "Locality Sensitive Hashing (LSH): The Illustrated Guide." Pinecone, [www.pinecone.io/learn/series/faiss/locality-sensitive-hashing/](http://www.pinecone.io/learn/series/faiss/locality-sensitive-hashing/). Accessed 4 Aug. 2023.

<sup>9</sup> Silver, David, et al. "Mastering the Game of Go with Deep Neural Networks and Tree Search." Nature, vol. 529, no. 7587, 2016, pp. 484-489, <https://doi.org/10.1038/nature16961>. Accessed 3 Aug. 2023.

**ELECTRICAL  
ENGINEERING  
TEAM PROJECTS**



# **Computational Modeling and Simulation of Inductive Traffic Loops for Vehicle Detection and Novel Classification Utilizing Deep Learning**

Lucas Philip, Sooraj Tharumia, Antony Tomy, Felix Yang

## **Abstract**

The effective management of vehicular traffic is of paramount importance in modern urban environments. In this context, due to their reliability and cost-effectiveness, inductive traffic loops have emerged as a widely adopted technology for vehicle detection, classification, and motional analysis. This research paper presents a comprehensive study of computational modeling and simulation techniques applied to inductive traffic loops. The primary objective of this study is to develop accurate and efficient computational models that can replicate the behavior of inductive traffic loops in various traffic scenarios. The paper introduces a novel approach combining electromagnetic theory and machine learning techniques to achieve this. By leveraging these interdisciplinary approaches, the proposed model can accurately predict the response of inductive traffic loops to passing vehicles, enabling precise vehicle detection and classification. Furthermore, the research explores the applications of these computational models in traffic management systems. Using discriminative artificial intelligence, emphasis is placed on developing advanced algorithms for real-time traffic analysis, such as traffic congestion detection. Overall, this research paper contributes to the advancement of traffic engineering and smart transportation systems by presenting a comprehensive computational framework for modeling and simulating inductive traffic loops. The proposed models and algorithms offer a reliable and cost-effective vehicle detection, classification, and motional analysis solution, ultimately paving the way for more efficient and intelligent traffic management systems.

## **I. Introduction**

### **A. Relevance**

Since its discovery in 1831 by legendary physicist Michael Faraday, electromagnetic induction has become the founding principle for much of modern technology. Power generators, electrical motors, and handheld electronic devices are just some examples of induction's near-limitless application. As the world continues to advance towards increasingly complex technologies and electronic systems, the principle of induction will only become more powerful and relevant in the global ecosystem.<sup>13</sup>

### **B. The Principle of Induction and Electromagnetism**

It has been known for centuries that electricity and magnetism are interconnected and may not be distinct at all. Faraday's experiments in 1831 discovered that whenever a conductor experiences a changing magnetic field, an EMF (electromotive force) is induced.<sup>1</sup> This EMF causes the movement of electrons in a conductor, thus generating what is conventionally referred to as 'electricity'. EMF functions on the underlying principle of magnetic flux ( $\phi_B$ ), which is a measure of the magnetic field in a conductor. Any change in  $\phi_B$  induces EMF, which generates an electrical current.<sup>1</sup> The variables governing  $\phi_B$  are the strength of the magnetic field  $B$ , the area  $A$  of the loop of wire perpendicular to the magnetic field, and the angle  $\theta$  between the magnetic field and the area vector of the loop. This relationship can be expressed as the following:

$$\phi_B = BA \cdot \cos(\theta) \quad (1)$$

It is important to note that  $B$  and  $A$  form a direct relationship with  $\phi_B$ ; that is, as the strength of the field or the area that the field intersects increases, the resulting magnetic flux will increase. The inverse is also true.

Equation 1 also demonstrates another important property of  $B$ ,  $A$ , and  $\theta$ : any change in any of those three variables causes a change in the value of  $\phi_B$ , thus inducing an EMF.

The value of the EMF ( $\varepsilon$ ) can be expressed as such:

$$\varepsilon = \frac{-Nd\phi_B}{dt} \quad (2)$$

where  $N$  indicates the number of loops of wire in a coil,  $\Delta\phi_B$  indicates the magnitude of the change in magnetic flux, and  $\Delta t$  indicates the time elapsed for the change in magnetic flux. Using this equation, we can quantify the EMF generated by any change in  $B$ ,  $A$ , or  $\theta$ .

Faraday's law and other major laws from Carl Friedrich Gauss and André-Marie Ampère of electromagnetism were recomposed to describe all of electromagnetism by James Clerk Maxwell.<sup>8</sup>

$$\oint_S (\vec{E} \circ \hat{n}) ds = \frac{q_{\text{enclosed}}}{\varepsilon_0} \quad (3)$$

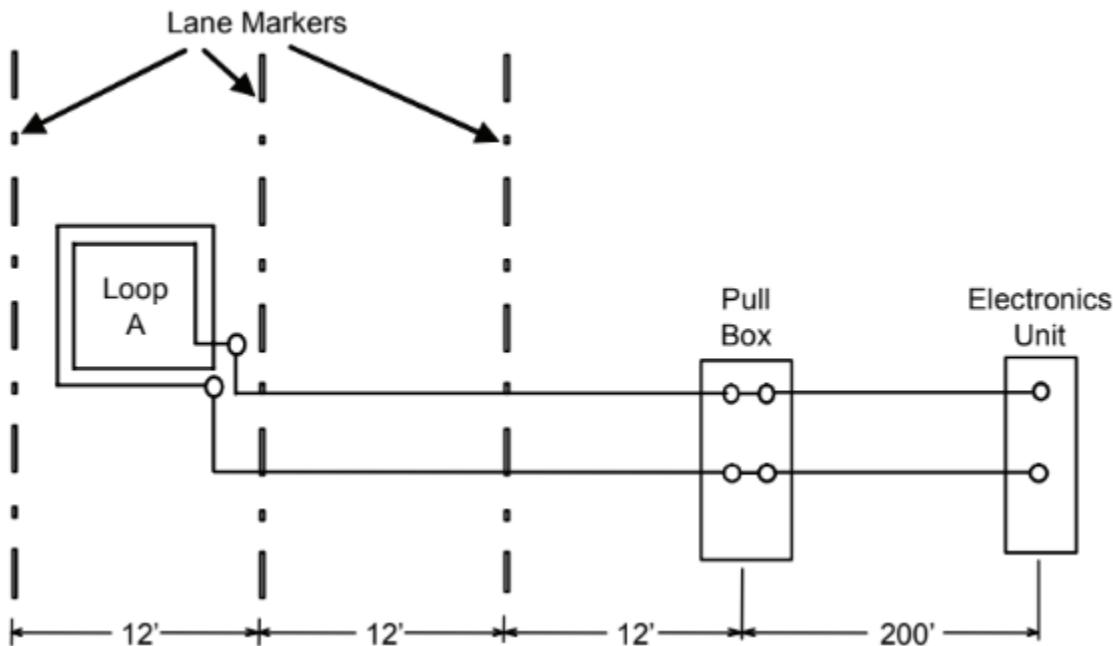
$$\oint_S (\vec{B} \circ \hat{n}) ds = 0 \quad (4)$$

$$\oint_C \vec{E} \circ d\vec{l} = -\frac{d}{dt} \int_S (\vec{B} \circ \hat{n}) ds \quad (5)$$

$$\oint_C \vec{E} \circ d\vec{l} = -\frac{d}{dt} \int_S (\vec{B} \circ \hat{n}) ds \quad (6)$$

These formalisms of electromagnetism were vital to not only constructing much of our modern technology but also allowing for the crucial computational simulations that can accurately model real-world systems.

### C. Inductive Traffic Loop Theory and Engineering



**Figure 1: Single inductive loop connected to loop box and an electronics unit.**

We can apply the principle of electromagnetic induction to traffic loops due to the metallic composition of vehicles. An induction loop placed below the ground will experience changes in inductance whenever the metallic undercarriage of a vehicle passes through, as the magnetic field generated by the metallic undercarriage interacts with the magnetic field generated by the induction loop. We term this interaction as an eddy current.<sup>13</sup>

Eddy currents are the result of an interaction between an electrically conducting surface and an active induction loop. Whenever a conductive surface passes through the magnetic field generated by an induction loop, eddy currents are induced in the conductive surface. These eddy currents generate an opposing magnetic field in the conductive surface, which in turn decreases the inductance in the induction loop.<sup>13</sup>

Induction traffic loops use this decrease in inductance to detect the presence of a passing vehicle. Whenever a vehicle passes over a buried induction traffic loop, an opposing magnetic field is generated by eddy currents, which decreases inductance in the loop and triggers sensors.<sup>1</sup>

This project did not explore the engineering of these sensors. Rather, we aimed to engineer the induction loop and investigate the interactions between induction loops and passing vehicles.

### D. Traffic Theory

Because inductive traffic loops aim to reduce traffic jams, it is important to first understand traffic flow. We may classify traffic into three distinct phases, as defined by Boris Kerner:<sup>5</sup>

- I. Free flow ( $F$ )
- II. Synchronized flow ( $S$ )
- III. Wide moving jam ( $J$ )

As additional notation, we define the flow rate of vehicles  $q$  and the vehicle density  $k$ , with maximum values  $q_{\max}$  and  $k_{\text{crit}}$ .<sup>12</sup>

The fundamental diagram of traffic flow gives the relationship between flow rate  $q$  and vehicle density  $k$ . We can already intuitively conclude that the more vehicles there are on a road, the slower they will be moving; that is,  $k$  and  $q$  express a direct relationship.<sup>12</sup>

In Kerner's theory, he describes two characteristics of S traffic, those being (i) non-interrupted rather than standstill and (ii) a tendency of synchronization of vehicle speeds, which is reflected in a relatively low probability of passing. Kerner describes J traffic as propagating through all other states of traffic flow (hence the term "moving jam") and reducing the speed of other states of traffic flow to the velocity of the jam front, which is either low or virtually zero.<sup>5</sup>

It is clear from the propagating nature of J traffic that the goal of traffic management is to avoid J traffic altogether by maintaining F or S traffic. It then becomes important to identify the source of J traffic, which Kerner describes as "bottlenecks". Sharp decreases in vehicle velocity (dubbed the 'breakdown phenomenon') and a decrease in highway capacity after the bottleneck (the "capacity drop") contribute to greater congestion after these bottlenecks.<sup>5</sup>

Although Kerner's observations were made predominantly on highway systems, the same concepts can be observed in urban and suburban road systems. Traffic lights are analogous to highway on-ramps in a sense; just as on-ramps cause a bottleneck effect on highways, inefficient traffic light systems cause a bottleneck effect on urban and suburban road systems. Yuan et al. describe that entry and exit flow volumes at urban intersections are largely dictated by the timing and efficiency of signal systems, which can be optimized through adaptive systems.<sup>15</sup> By using inductive traffic loops to determine the degree of vehicle volume at intersections, we can use algorithms to determine the optimal signal timing to reduce or prevent congestion.

## E. Conductivity and Resistivity

Electrical conductivity is an extensive property of materials. The quantum basis for conductivity lies in an atom's ability to hold electrons in certain precise energy levels, beginning from low energy levels. We define the highest energy level to which electrons fill as the Fermi level; the position of this Fermi level is incredibly relevant to the conductive properties of an element, as only electrons above this level can be transferred. Generally, metals have comparatively many electron levels near this Fermi level; thus, metals have a greater ability to transfer electrons, making them more conductive.<sup>7</sup>

We must also consider resistance and resistivity when we analyze the movement of electrons through metals. Although similar, the distinction between the two is that resistivity is an intrinsic property (i.e. constant throughout the conductor, regardless of shape), whereas resistance is dependent on the shape of the conductor. We can express the resistance  $R$  of a conductor as follows:<sup>7</sup>

$$R = \rho \frac{L}{A} \quad (7)$$

where  $\rho$  is the electrical resistivity of the conductor,  $L$  is the length of the conductor, and  $A$  is the cross-sectional area of the conductor (Note: the units of  $\rho$  is ohms/meter, the units of  $L$  is meters, and the units of  $A$  is meters<sup>2</sup>, thus the units of  $R$  is ohms).

Note that we can alternatively define resistivity  $\rho$  as the relationship between the magnitude of the electric field  $E$  at a particular point and the magnitude of the current density  $J$  at the same point:<sup>7</sup>

$$\rho = \frac{E}{J} \quad (8)$$

We define current density as the amount of current flowing per unit area of the wire. The equation for current density shows that as the cross-sectional area of the conductor decreases, the current density increases. Thus, to maximize the current density and minimize resistance, thin metal wires with a high conductivity are optimal conductors.<sup>7</sup>

The use of copper is ideal in induction traffic loops because of its high conductivity and low resistance. Additionally, it is abundant and easy to source. The applications of copper will be further illustrated in Part II.

## F. Applying Ansys High-Frequency Structure Simulator

In order to be able to simulate the inductive properties of induction traffic loops, Ansys High-Frequency Simulation Software (HFSS) was used in order to simulate the loop and vehicle traveling through the loop and the inductive properties/changes going through the simulated loop, respectively.

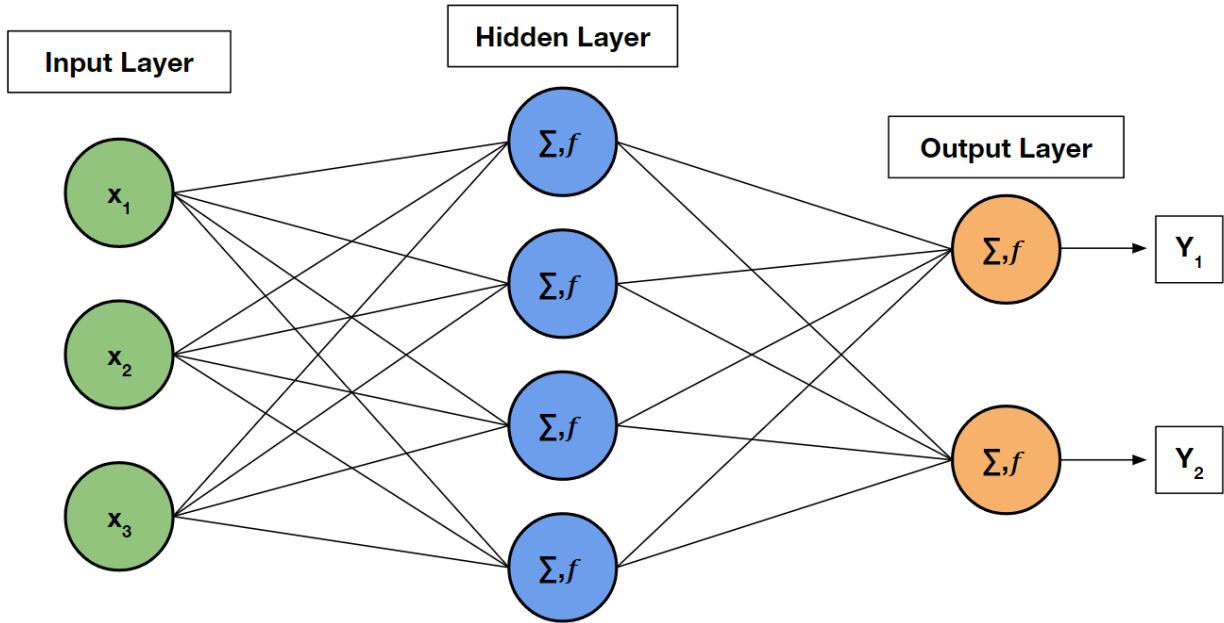
Ansys HFSS is a 3D electromagnetic simulation software that is typically used for antenna design and high-frequency simulation. With this software, we can accurately model and simulate the interactions between buried multi-turn induction loops and vehicles with different dimensions and properties. More specifically, we can simulate the decrease in inductance whenever a vehicle passes over the loop (the ‘signal’ mechanism), which is caused by the generation of eddy currents. For the purposes of our project, we model the interactions between the multi-turn copper wire loops and the undercarriages of vehicles using HFSS’s built-in boundary creation, meshing, and excitation functions which will be further discussed in the methodology.

Ansys HFSS’s various environments and tools give us a versatile method to experiment with different vehicle shapes, materials, and other properties. Since the decreases in inductance will largely depend on the vehicle’s properties, we can classify the vehicles based on the magnitude of this decrease.

## G. Classification using Neural Networks

We can determine the type of vehicle passing over an induction loop by analyzing the magnitude of the decrease in induction that the vehicle causes. In order to perform this classification task, we implemented a multilayered neural network.

Neural networks function much like the neurons in the human brain. In the process of learning, we reinforce specific neuron-neuron connections by observing patterns. In neural networks, vectors of weights are adjusted to reflect patterns that a neural network may observe in a dataset.<sup>2</sup>



**Figure 2: A neural network visualized**

The ‘learning mechanism’ of neural networks is what makes them incredibly powerful. As a neural network goes through epochs, or generations, of learning, they become increasingly accurate, often surpassing human intuition and observation.<sup>2</sup>

Fundamentally, neural networks contain three components. The first component is the input layer, which accepts a vector of features, or inputs, to be processed by the network. The second component is the hidden layers, which perform computations on the given vector input and adjusts its weights accordingly. The final component is the output layer. In classification tasks, the output layer determines the classification of each event.<sup>2</sup>

Each layer is composed of perceptrons. The function of a perception is essentially to perform computations and can be summed in three steps:

1. Application of a weight vector  $w_i$  to a given input vector
2. Addition of a bias  $b$
3. Application of an activation function  $f(x)$

We can also write this function as a mathematical expression:

$$f(\sum_{i=0}^n w_i x_i + b) \quad (9)$$

where,  $f(x)$ ,  $b$ , and  $w_i$  are as described above.  $x_i$  indicates an input vector containing  $n$  total elements (hence the summation from  $i = 0$  to  $n$ ). Neural networks learn by adjusting  $w_i$  every generation to better model the patterns of the dataset. The adjustment of  $w_i$  is analogous to the system of strengthening certain neural connections in the human learning process.<sup>2</sup>

In order to train a neural network, a sufficient dataset is required for the model to train on. There are multiple common strategies used for learning, one of which is supervised learning. In supervised learning, the model is given both the features and the labels in a dataset.<sup>2</sup>

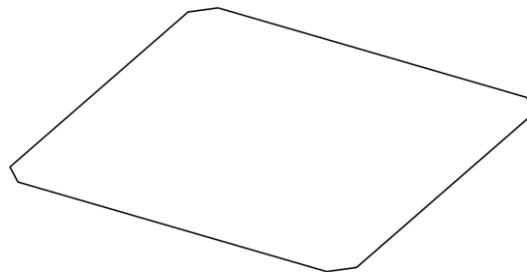
## II. Methodology

### A. Simulation in Ansys HFSS

#### 1. Designing an Induction Loop in Computer-Aided Design (CAD)

In order to simulate the effects of the movement of vehicles across an inductive loop, an appropriately sized and designed inductive loop must be made. To be able to design the loop to the greatest ability, the CAD software Onshape was used in this process. By using this software, and not the “Drawing” functions given in Ansys HFSS, we were able to make the loop much more accurate to the real-world sizes and design the loop much faster.

Following the guidelines of the United States Department of Transportation traffic handbook, an octagonal-shaped inductive loop was created with the dimensions 1.8 meters x 1.8 meters.<sup>11</sup> The diameter/width was taken into account, as well, having a size of 1.628 mm. These values were provided by the US DoT traffic handbook, so a similar loop was designed to model the types of inductive loops that can be found in the real world.<sup>11</sup>



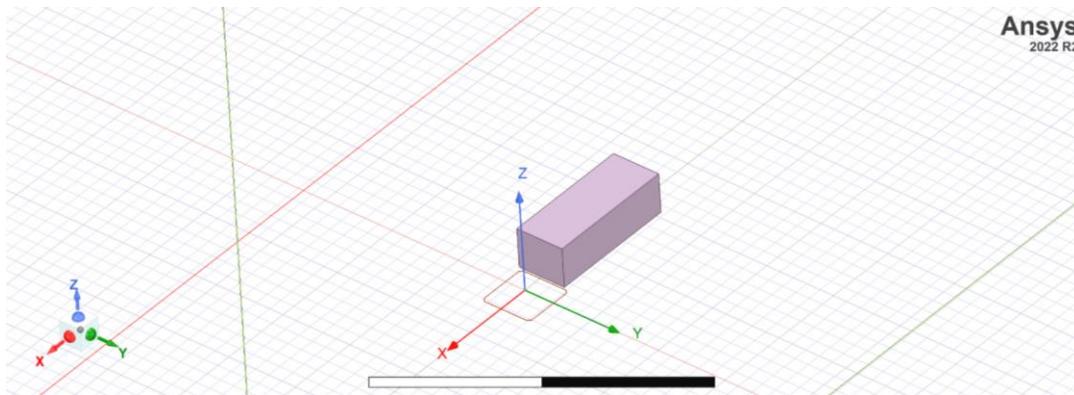
**Figure 3: Octagonal one-turn loop used for simulations, modeling, and data collection**

#### 2. Designing an Induction Traffic Loop Environment in HFSS

As Section I describes, Ansys HFSS is a powerful software engine that can generate environments for simulating electromagnetic interactions. In addition to its ability to simulate electromagnetic interactions, Ansys HFSS also provides various tools for object design, including but not limited to dimensional manipulations, materials, transformations, and more.

In order to simulate an induction traffic loop, we modeled the inductive loop lying underneath the ground with copper and the vehicle with stainless steel. The vehicle travels above the ground with specific ground clearance, dependent on the type of vehicle taken into account. The eight-sided shape additionally provides more detection points on the exterior of the loop, thereby reducing the risk of false readings. By making the vehicle and copper conductive metals, electromagnetic induction in the inductive loop can be harnessed to classify what kind of vehicle it is.

### 3. Designing Classes of Vehicles



**Figure 4: Depiction of the simulation setup in Ansys HFSS.**

Table 2. Dimensions of Different Vehicles

Type	Model	Length (mm)	Width (mm)	Height (mm)	Ground Clearance (mm)
Bicycle	Single Speed	1788	42	1080	295
Bicycle	Fat Tire	1955	101.6	1092	330
Motorcycle	Pulsar RS200	1999	765	1114	152
Motorcycle	Honda CB1000R	2105	755	1095	130
Sedan	Honda Civic	4450	1695	1385	155
Sedan	Mitsubishi Lancer	4625	1760	1490	165
SUV	Isuzu Mu7	4955	1800	1805	225
SUV	Audi Q7	5086	1983	1737	205
Truck	Toyota Tundra	5814	2029	1935	269
Truck	The Monterra 4x4	6394	2438	2845	559
18-Wheeler	Cerasis 2015	12912	2540	3962	550
Bus	Sorban 9.5	9600	2525	2950	254
Bus	Daewoo FX212	12000	2490	3550	254

**Figure 5: Vehicle dimensions.**

As shown in Figure 4, we modeled different classes of vehicles as rectangular prisms. Each vehicle was assigned dimensions based on real-world designs. The length, width, height, and ground clearance were taken into account while designing these vehicles.

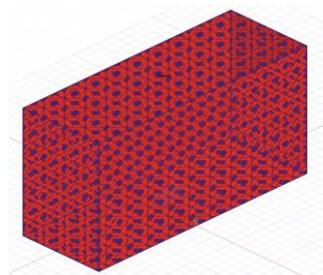
The first class we considered was Class 0 vehicles, which consist of motorcycles/bicycles. The dimensions of our model Class 0 vehicle were  $2.1 \times .75 \times 1.2 \times 0.13$  meters. The next class we considered was Class 1 vehicles, which consisted of Sedans and SUVs. The dimensions of our model Class 1 vehicle were  $4.5 \times 1.8 \times 1.33 \times 0.2$  meters. The final class that we considered was Class 2 vehicles, which consisted of Pickup Trucks. The dimensions of our model Class 2 vehicle were  $5.75 \times 2 \times 1.9 \times 0.26$  meters.

These distinct classes give us the ability to classify these vehicles based on reasonable differences in the data, which in this case is the instantaneous change in magnetic field from the induced current dB.

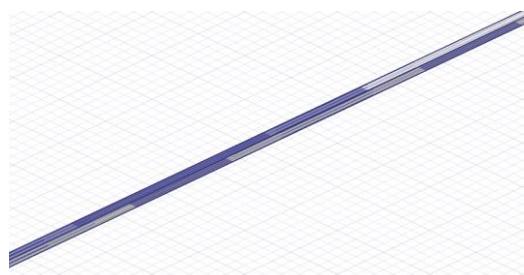
#### 4. Setting up Ansys HFSS Options

Prior to running the simulation, the appropriate options and parameters must be set to correctly obtain the data. The necessary setup for the simulation was found using the “Validation Check” and the parameters in this check. These remaining steps included the following: Boundaries, Ports, Mesh Operations, Solution Setup, and Optimetrics. These steps will be outlined throughout this section.

Two main boundaries were used in this simulation setup. The main boundary used in this project was the Open Region, which incorporated a radiating region. Within this region, the electromagnetic fields released during the simulation can propagate and get excited at the appropriate operating frequency, which in this project was 20 kHz. The region is rather large compared to the loop/vehicle combination, so electromagnetic waves can propagate rather well. The next boundary used in this project was the PerfE boundary, which is expanded as a Perfect E-Field boundary. This was placed on the loop directly and made a region for a perfect electric field coming off of the loop, meaning that an electric field will propagate from the loop perpendicular to the surface at all areas. This maximizes the electromagnetic induction in the loop due to the reduced possible angle between the magnetic field vector and the area vector as the vehicle moves over the loop,  $\theta$ . These two boundaries can be seen in Figures 6 and 7, respectively.

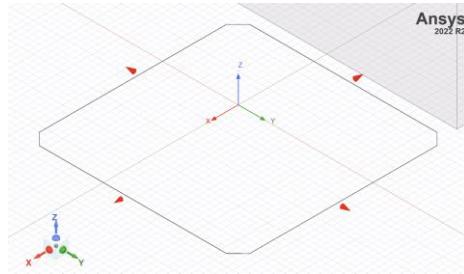


**Figure 6: The radiating region**

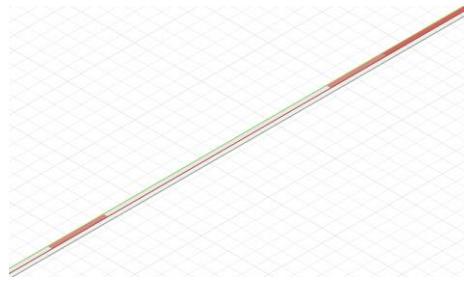


**Figure 7: Perfect E boundary**

A lumped port was assigned to the top surface of the loop in this project. The lumped port is essentially where the electromagnetic fields emerge from and is a critical part of this project. A lumped port was applied in this project as it needs to be inside the radiating region in order for the electromagnetic fields to interact with the vehicle and the rest of the loop. As seen in Figures 8 and 9 show that the lumped port can be seen on the loop, and the direction that the electromagnetic fields will propagate.

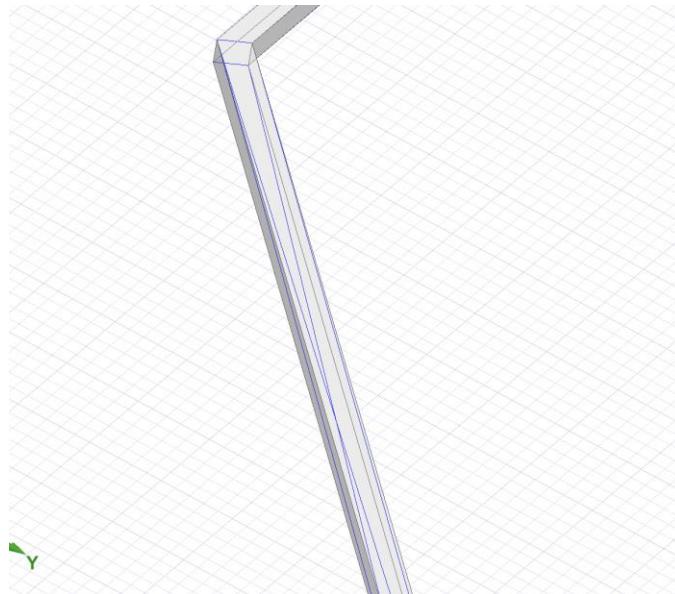


**Figure 8: Direction the electromagnetic waves propagate**



**Figure 9: The lumped port**

An on-selection, length-based mesh operation was placed on the top surface of the loop. The mesh operation is responsible for computing the changes in the electric and magnetic fields, which in turn, makes the simulation much more accurate. The mesh length was set to 1.65 meters in this project, which is slightly smaller than the overall size of the actual loop. The small triangulations can be seen on the loop (as seen in Figure 10), and these are critical to carrying the computational might needed for quantifying these changes in the magnetic and electric fields.



**Figure 10: Mesh operation on the loop.**

A frequency sweep was used in this project from 1 kHz to 20 kHz at a fast sweep type. The fast sweep type was used in order to reduce simulation time and acquire the general basic trend needed in the data. This general trend was used by the neural network algorithm to classify the type of vehicle going over the loop. The frequency sweep is used to gather the data by changing the propagation frequency of the electromagnetic waves (from 1 kHz to 20 kHz). By using this sweep, the instantaneous change in magnetic field dB can be calculated at different frequencies.

A parametric optometric was used to move the vehicle across the loop. In order to do this, a variable named corner was set as the bottom rightmost corner of the box, and with this variable, we were able to move the box by changing the variable value. The parametric settings in Figure 11 were applied to the truck. At each step, a frequency sweep was run in order to collect the data as the vehicle moved across the loop. The change in magnetic field was tracked as the vehicle moved across the loop, which can be plotted and compared for each different class of vehicle.

Sync #	Variable	Description
	corner	Linear Step from -3500mm to 5000mm, step=500mm

**Figure 11: Parametric sweep assignment for the corner variable.**

## B. Utilizing an Artificial Neural Network Algorithm for Classification

### 1. Data Preparation and Preprocessing

After the simulations of the inductor traffic loops were run for each of the vehicle classes, the electronic data of the copper loop during each step of the sweep was exported into a comma-separated values (CSV)

format during each trial. Google Cloud computational resources were incorporated in order to accelerate the machine learning process in this project through Google Colaboratory. This data was then preprocessed in the Python 3 programming language. Then, we labeled each event with its correct index, with 0 representing Class 0 vehicles (i.e. motorcycles), 1 representing Class 1 vehicles (i.e. sedan-like vehicles), and 2 representing Class 2 vehicles (i.e. pickup trucks and larger vehicles).<sup>1,11</sup> The input data would contain 7 features from the 7 steps of the sweep that were chosen from each of 30 training examples from the Ansys HFSS. Upon labeling the data and separating it into electronic data as the input and the vehicle class designation as the output, the train-test-split function from the sklearn library was implemented in order to randomly allocate 20% of the data to testing, leaving the rest 80% for training of the neural network model.

We defined our features as the positions at which dB readings were taken in the parametric sweep. Although the positions we considered in each event were not constant, we aimed to model the general pattern of the data through the features we chose. We designated each of these position features as  $L_1$  through  $L_7$ , and we designated the label of each event as the vehicle class to which the vehicle in the event belonged to.

## 2. TensorFlow / Keras API

TensorFlow is an open-source machine learning and neural network framework released by Google in 2015 for complex numerical computations and machine learning tasks. TensorFlow offers a high-level API called Keras, which we used in our classification task. Keras enables developers to create dataflow graphs through object-oriented programming in order to represent their machine learning models, such as artificial neural networks (ANN), in an optimizable and efficient way. The library is very compatible with the Python programming language, which additionally gives developers the functionality of traditional Python programming. The latest 2.12.0 version of TensorFlow, along with the 2.13.1 version of Keras, was employed to develop the ANN algorithm.<sup>10</sup>

## 3. Model Architecture

Our initial architecture was composed of the input layer followed by two fully-connected hidden layers of 64 neurons each and a fully-connected output layer with 3 neurons. (Note: TensorFlow allows us to specify the shape of our input vector as a parameter in our first hidden layer. Since we aren't performing any pre-training functions, we don't need a designated input layer<sup>10</sup>). We used the ReLU activation function (Rectified Linear Unit) in each hidden layer and the softmax activation function in the output layer.

```
model = keras.Sequential()
model.add(layers.Dense(7, activation="relu"))
model.add(layers.Dense(64, activation="relu"))
model.add(layers.Dense(64, activation="relu"))
model.add(layers.Dense(3, activation="softmax"))
```

**Figure 12: Implementation of a Sequential NN using TensorFlow**

---

<sup>1</sup> Class 0, 1, and 2 as defined in this paper are roughly analogous to the US DOT's vehicle classification system. Class 0 is analogous to DOT Class 1, Class 1 is analogous to DOT Class 2, and Class 2 is analogous to class 3 vehicles.

The purpose of an activation function is to define the output of a node, given a set of inputs. Activation functions, like ReLU and softmax, introduce non-linearity into a neural network. Non-linearity allows a neural network to perform more complex classification tasks with a greater degree of accuracy.<sup>2</sup>

Figures 13 and 14 are graphical representations of the ReLU and softmax activation functions. The ReLU activation function,  $\max(0, x)$ , returns an output  $x$  if the input  $x$  is positive and zero otherwise. ReLU is highly advantageous as an activation function because of its simplicity; other activation functions, like the Sigmoid and tanh activation functions, involve much more complex mathematical operations, which is less efficient and computationally intensive.<sup>2</sup>

The softmax activation function turns a vector  $z$  of  $k$  values into a vector  $v$  of  $k$  values, where  $\sum_{j=1}^k z_j = 1$ . In essence, the softmax function transformation  $v_0$  to be a normalized probability distribution. In cases where the inputs to the penultimate layer of a neural network are not scaled, using the softmax function makes the output vector much more interpretable and easier to work with.<sup>14</sup>

The softmax function is more complex than the ReLU function, and can be represented as such:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (10)$$

where  $z_i$  is the set of all elements of the input vector. The summation function in the denominator ensures that the output values will sum to 1 and each individual element is in the range (0, 1). This gives us a valid probability distribution as our output vector.<sup>14</sup>

The Keras API applies the specified activation function to the input vector at each node. This added non-linearity to our classification network, which was appropriate for our non-linear classification problem.

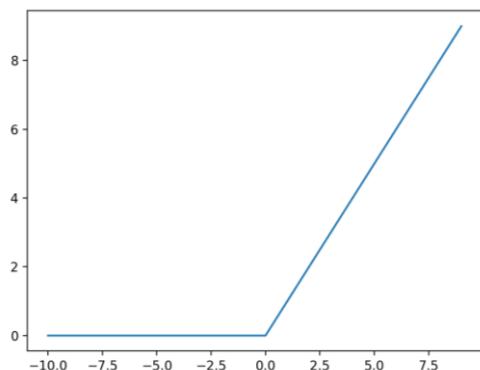


Figure 13: ReLU activation function

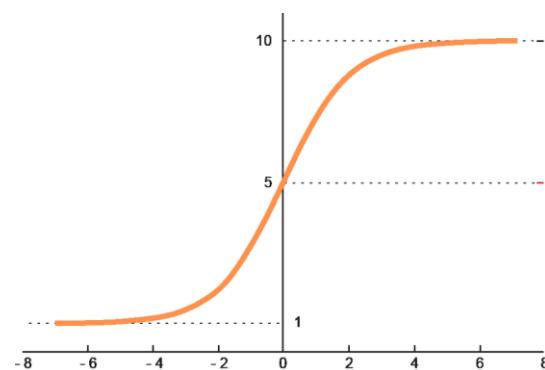


Figure 14: Softmax activation function

```
model.compile(loss = "categorical_crossentropy", optimizer = "adam", metrics=["accuracy"])
```

Figure 15: Compiling the neural network

Note that TensorFlow Keras API allows us to specify an optimizer and a loss function when we compile our model. An optimizer is an algorithm that adjusts the attributes of the neural network to minimize the loss function. A loss function is, in essence, a measure of how inaccurate the model is. The loss function is what the model attempts to minimize every generation.<sup>2</sup>

#### 4. Neural Network Training Strategies

We can separate the process of training the neural network into three components as follows:

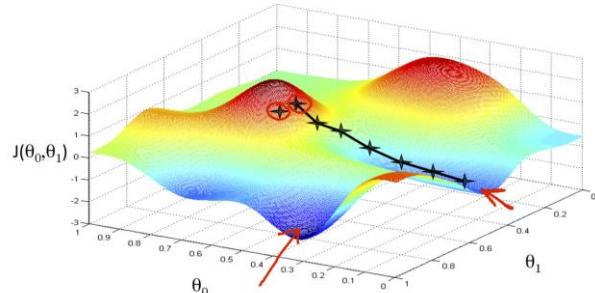
1. Forward propagation
2. Back propagation
3. Manual adjustments of hyperparameters

In every generation, or epoch, the neural network performs forward and back propagation. Forward propagation follows the process described in I(G), where an input vector is multiplied by a weight vector at every perceptron, and the sum of the products is fed into an activation function. After the predicted label and the model's confidence in the predicted label is given, the model checks how inaccurate its prediction was when compared to the true label, and the model adjusts the weight vectors and bias term accordingly. This optimization process of changing weights and biases is called back propagation. Finally, we also manually adjusted the number of hidden layers, the number of perceptrons in each hidden layer, and the activation functions used in each layer. These qualities of the neural network are called hyperparameters, and we have to manually adjust these values to better fit the data we want to model.<sup>2</sup>

In the process of optimization, there are numerous strategies that a neural network can take. In our model, we used a stochastic gradient descent optimizer, Adam, to adjust the weights and biases of our neural network.

##### Gradient descent algorithm

```
repeat until convergence {
     $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ 
    (for  $j = 1$  and  $j = 0$ )
}
```



**Figure 16: Mathematical expression of a gradient descent algorithm**

**Figure 17: Graphical representation of a gradient descent plot.**

A stochastic gradient descent optimizer attempts to minimize the loss function as much as possible. In every iteration, the optimizer calculates the slope of the objective function (a loss function) with respect to each feature that we train the model on. Then, the optimizer calculates a 'step size' and applies this step size to the weight vector. In the next generation, the optimizer performs the same process and continues to adjust the weight vectors until, ultimately, the gradient is as close as possible to 0.<sup>3</sup>

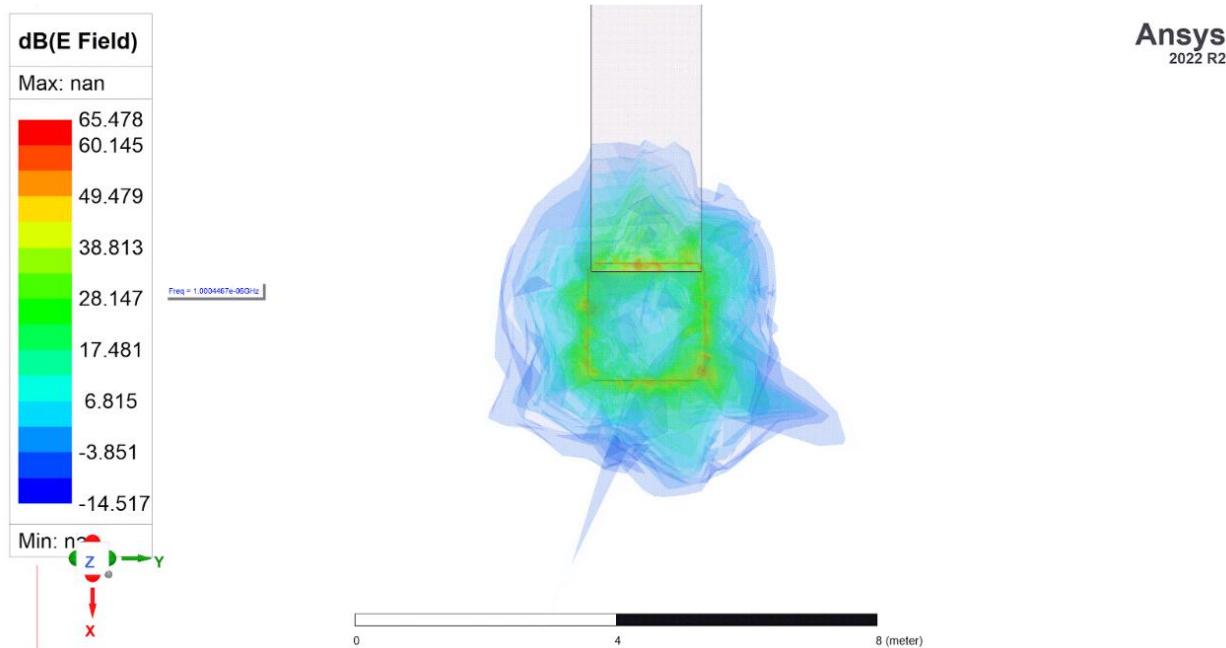
The Adam optimizer adjusts weight and bias vectors by this strategy. But in addition to the optimizer, we also needed to define a loss function for our model. For our model, we used the sparse categorical cross-entropy loss function, which follows the mathematical expression shown in Figure 18 below.<sup>6</sup>

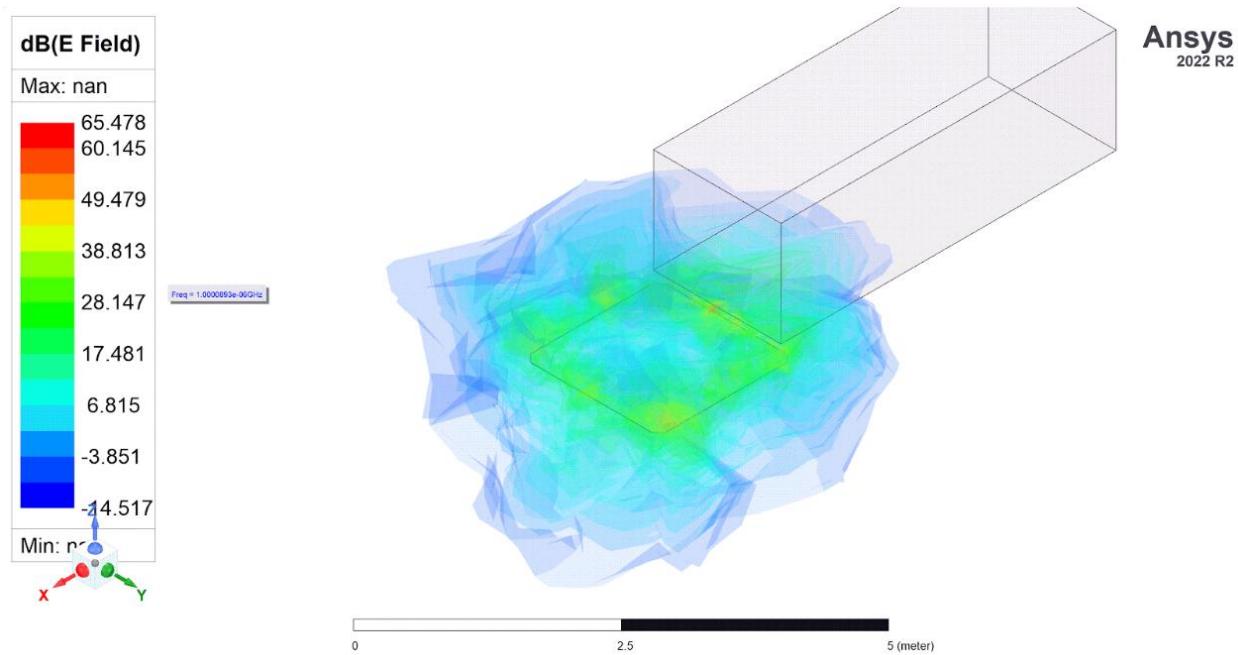
$$\text{Loss} = - \sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

Multiclass Cross-Entropy

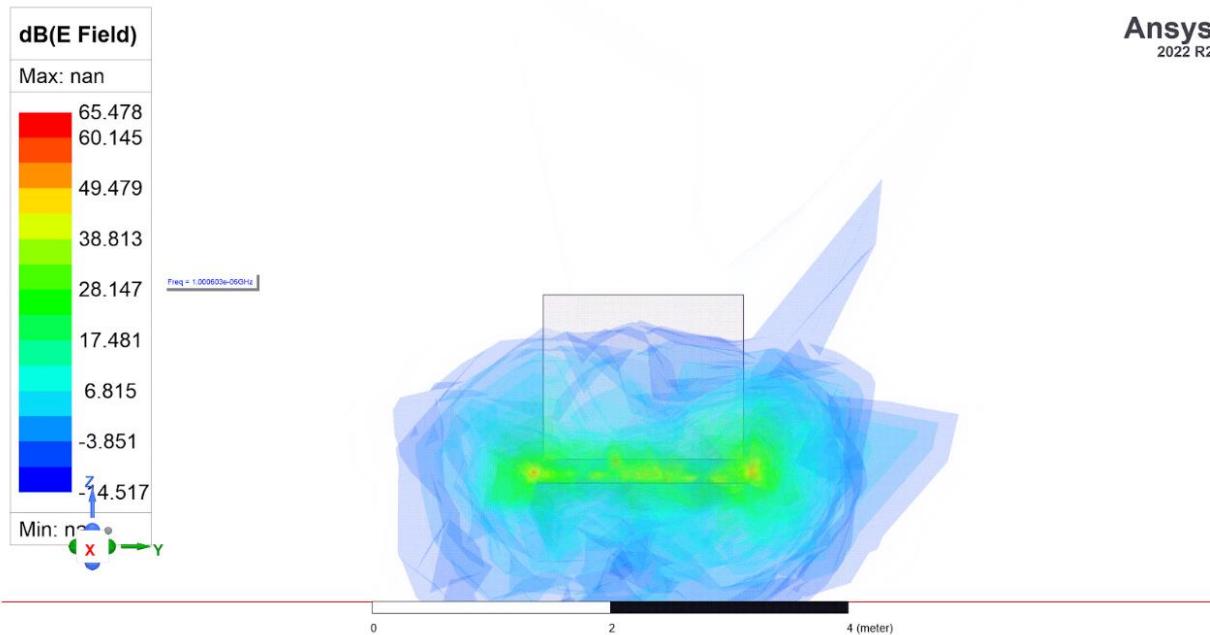
**Figure 18. Categorical Cross-Entropy Loss Function**

Because our classification task was multi-class, we needed to implement a categorical loss function that allows for more than binary labeling. Generally, as the difference between our predicted label and the true label increases, our calculated loss increases. Thus, whenever the model is extremely inaccurate, the magnitude of the step size (as defined in the optimizer) increases.<sup>9</sup>

**Figure 19: Bird's eye view of the simulation showing the dB values of the magnetic field in the loop placed beneath the sedan model.**



**Figure 20:** Three-dimensional view of the simulation showing the dB values of the magnetic field in the loop placed beneath the sedan model.



**Figure 21:** Front side view of the simulation showing the dB values of the magnetic field in the loop placed beneath the sedan model.

### III. Results

#### A. HFSS Simulation Results

In the above figures, we can see that the magnetic field generated by the simulation setup is reflective of a real-world situation. In Figure 19, the birds-eye view of the inductive loop shows the magnetic field radiating outward from the loop. In Figure 21, the model also represents how the magnetic field radiates upwards and downwards from the loop. This aspect of the simulation is important because it accurately depicts how the loop penetrates through the concrete road to reach the vehicle's chassis.

Each class had 10 different vehicles—10 sedans, 10 motorcycles, and 10 trucks—to gather a large amount of data. Each vehicle modeled the dimensions of a true vehicle in the real world. At each position, L1 - L7, the instantaneous change in magnetic field dB was found by the Ansys simulation. The reason L1 - L7 was used instead of discrete position values is because the size of the classes is vastly different—for example, a motorcycle is much smaller than a truck. Because of these size differences, a discrete value of a position might mean something different for each class. To deal with this, the L1 - L7 was used such that it is a position relative to the loop. For example, L1 is at the starting point on one side of the loop, L4 is a position where the vehicle is in the middle of the loop, and L7 is a position on the outside of the loop on the other side. The data found at each of these positions can be found below in Figure 22.

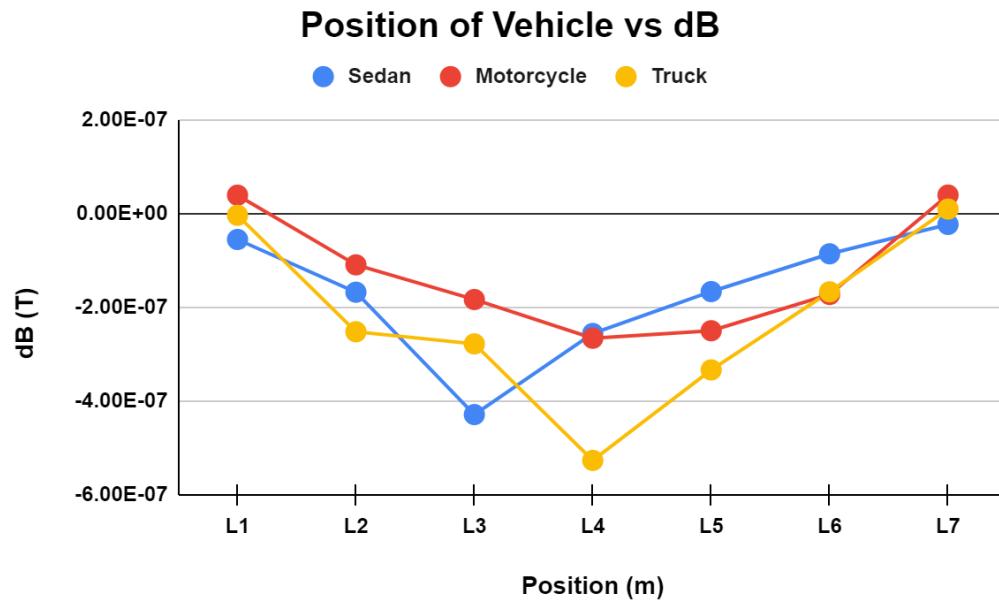
L1	L2	L3	L4	L5	L6	L7	LABEL
2.53E-09	-1.88E-07	-2.94E-07	-2.34E-07	-1.66E-07	8.13E-10	5.97E-08	sedan
-2.28E-09	-1.52E-07	-2.53E-07	-1.94E-07	-1.65E-07	-1.14E-07	-1.14E-07	sedan
-2.50E-07	-2.66E-07	-3.40E-07	-4.13E-07	-4.70E-08	-9.86E-08	-7.94E-08	sedan
5.44E-08	-5.57E-08	-4.00E-07	-1.59E-07	-1.76E-08	-1.47E-08	-1.18E-08	sedan
-2.18E-07	-3.21E-07	-1.41E-06	-2.72E-07	-2.58E-07	-2.60E-07	1.46E-07	sedan
-1.61E-07	-2.19E-07	-5.18E-07	-2.86E-07	7.14E-08	3.77E-08	-1.44E-07	sedan
7.21E-08	-1.71E-08	-2.69E-07	-1.16E-07	-1.22E-07	-9.63E-09	1.02E-08	sedan
-2.80E-08	-2.53E-07	-2.91E-07	-1.91E-07	-1.20E-07	-3.59E-08	-1.09E-07	sedan
1.91E-08	4.62E-08	-3.05E-07	4.81E-08	-3.03E-07	-1.89E-07	-8.67E-08	sedan
-4.26E-08	-2.53E-07	-2.08E-07	-7.41E-07	-5.38E-07	-1.76E-07	9.72E-08	sedan
-4.21E-08	-5.13E-08	-1.48E-07	-1.96E-07	-3.05E-07	-2.01E-07	-5.54E-08	motorcycle
1.86E-07	-1.98E-07	-3.63E-07	-3.47E-07	-3.47E-07	-1.38E-07	1.22E-07	motorcycle
-9.65E-09	-1.05E-07	-2.20E-07	-3.91E-07	-3.86E-07	-2.40E-07	1.04E-07	motorcycle
1.84E-07	-1.17E-07	-1.94E-07	-3.27E-07	-3.59E-07	-4.08E-07	6.39E-08	motorcycle
-1.24E-07	-1.28E-07	-1.55E-07	-2.50E-07	-2.57E-07	-1.26E-07	-4.84E-08	motorcycle
1.84E-07	-1.17E-07	-1.94E-07	-3.27E-07	-3.59E-07	-4.08E-07	6.39E-08	motorcycle
-6.28E-08	-2.87E-08	-5.37E-09	-2.64E-08	-2.13E-07	-2.19E-08	-8.82E-09	motorcycle
-5.92E-09	-1.33E-07	-2.27E-07	-2.66E-07	-9.82E-08	-3.79E-08	3.66E-08	motorcycle
1.28E-07	-5.34E-08	-1.23E-07	-2.06E-07	-1.23E-07	-1.38E-07	2.49E-08	motorcycle
-4.70E-08	-1.64E-07	-2.05E-07	-3.24E-07	-5.25E-08	-7.19E-09	9.09E-08	motorcycle
-3.02E-08	-2.76E-07	-9.52E-08	-7.37E-07	-5.02E-07	-3.78E-07	-6.16E-08	truck
-6.77E-08	-3.49E-07	-1.03E-07	-4.48E-07	-2.58E-07	-8.05E-08	7.77E-08	truck
-9.00E-09	-2.53E-07	-3.62E-07	-2.26E-07	-2.26E-07	-9.55E-08	4.93E-08	truck
-5.32E-08	-3.04E-07	-4.00E-07	-3.48E-07	-2.67E-07	-5.76E-08	2.15E-07	truck
3.18E-07	-2.56E-07	-2.28E-07	-5.82E-07	-3.46E-07	-2.90E-07	-2.43E-07	truck
8.75E-08	-9.51E-08	-1.06E-07	-4.51E-07	-2.40E-07	9.80E-08	2.51E-08	truck
7.82E-08	-2.10E-07	-4.25E-07	-6.13E-07	-2.86E-07	-4.69E-08	2.30E-08	truck
-7.37E-08	-2.46E-07	-3.45E-07	-8.04E-07	-3.33E-07	-3.03E-07	4.32E-08	truck
-2.15E-07	-2.69E-07	-3.54E-07	-6.26E-07	-5.36E-07	-3.17E-07	-2.47E-07	truck
-7.28E-08	-2.64E-07	-3.62E-07	-4.30E-07	-3.41E-07	-2.00E-07	2.13E-07	truck

**Figure 22: Thirty example events (10 sedan, 10 motorcycle, 10 truck) from the training set**

The average dB at each of these positions can be seen below in Figure 23. Additionally, the graph of position vs. dB can be seen in Figure 24.

Sedan Avg						
L1	L2	L3	L4	L5	L6	L7
-5.54E-08	-1.68E-07	-4.49E-07	-2.56E-07	-1.67E-07	-8.59E-08	-2.32E-08
Motorcycle Avg						
L1	L2	L3	L4	L5	L6	L7
3.91E-08	-1.10E-07	-1.83E-07	-2.66E-07	-2.50E-07	-1.73E-07	3.93E-08
Truck Avg						
L1	L2	L3	L4	L5	L6	L7
-3.79E-09	-2.52E-07	-2.78E-07	-5.27E-07	-3.34E-07	-1.67E-07	9.47E-09

**Figure 23: Table of average dB readings at all positions L1 through L7**

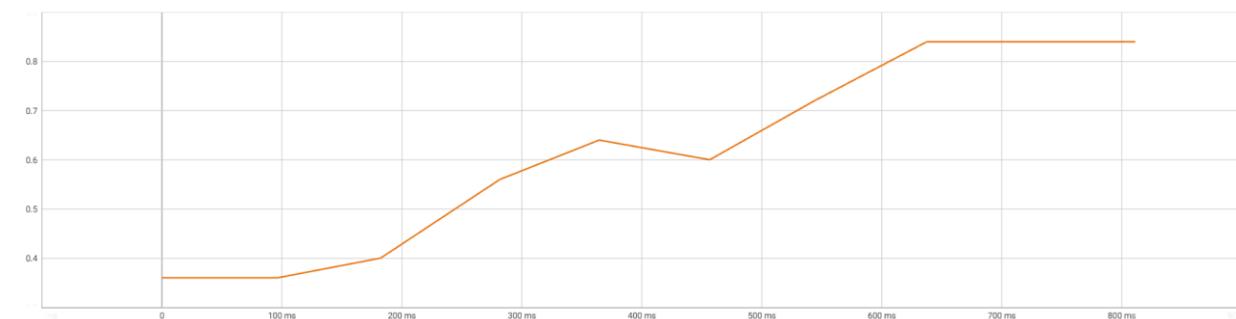


**Figure 24: Graph of dB vs. position**

As seen in the graph above, the truck had the greatest change in the magnetic field, and the motorcycle had the lowest change in the magnetic field. This is reflective of what would be found in the real world because a truck is so much bigger than a motorcycle, so there is more area for the undercarriage of the vehicle to interact with the loop and induce a greater magnitude current in the loop. This greater magnitude current in the loop will result in a greater magnetic field by the Law of Biot-Savart.<sup>1, 8, 13</sup>

## B. Neural Network Evaluation Metrics

In our project, we trained the neural network over 10 epochs. The results of training are shown in Figures 25 and 26, where we can see that over these epochs, the neural network was able to increase its accuracy to 88% and decrease the loss function significantly.



**Figure 25: Accuracy vs Time Graph in Tensorboard**

**Figure 26: Loss vs Time Graph in Tensorboard**

	precision	recall	f1-score	support
0	0.88	0.78	0.82	9
1	0.82	0.90	0.86	10
2	1.00	1.00	1.00	6
accuracy			0.88	25
macro avg	0.90	0.89	0.89	25
weighted avg	0.88	0.88	0.88	25

**Figure 27: Final Evaluation Metrics of ANN model**

If we consider these evaluation metrics more in-depth, we can deduce from the precision score of 1.00 that for Class 2 (trucks), the model predicted no false positives. The model also achieved a recall score of 1.00 for Class 2, indicating that it correctly identified all the truck events in the dataset. Although the precision and recall scores for Class 0 and 1 vehicles were not perfect, they were still very high, which indicates that our model was able to correctly classify the type of vehicle in most cases.

## IV. Analysis and Limitations

### A. Analysis and Limitations of the Neural Network

In training our neural network, it is important to note that in relative sizes, our dataset was very small. We were greatly limited by computational power and time, as it took a significant amount of time to perform each simulation (>10 minutes) in Ansys HFSS. We attempted to increase the efficiency of our simulations by lowering the fineness of our meshes and using different sweeps, but obtaining usable simulation data in a feasible amount of time still proved to be a challenge. Because we trained on a relatively small dataset, there may be trends that we may only have been able to observe with a larger amount of data; thus, it may be possible that our model may not be able to correctly classify new events.

There is also the possibility of our model overfitting the data. In machine learning, overfitting is described as a model learning each specific data point in the training set, rather than learning to recognize general patterns. This arises in the case of a small dataset with a high variance. Because our dataset was small, it is possible that overfitting may have impacted our model's ability to correctly classify the test data.<sup>2</sup>

Due to a lack of computational power, the neural network could not be trained for further training iterations or epochs. This could have hindered further convergence of the Adam gradient descent algorithm in converging on the minimum of the cost function, thereby providing a less accurate model.

## B. Analysis and Limitations of Ansys HFSS Software

Although Ansys HFSS is an incredibly powerful software, we encountered many issues with time efficiency. Many factors impact the efficiency of an HFSS simulation, including the mesh, number of steps in a parametric, complexity of objects, and so on. The detail and comprehensiveness to which a simulation is performed greatly impacts the time needed to run the simulation, and with limited computational resources, we chose to sacrifice some detail in order to run more simulations and collect a dataset.

## V. Future Research

There are multiple avenues that could be explored further as extensions in our research.

### 1. Expanding the Neural Network Architecture

The architecture of our neural network was also relatively simple. In consideration of computational runtime, we chose to include two hidden layers. However, if we were to use many more layers (on the order of tens or hundreds), we would be able to introduce more complexity into our model, which may contribute to a higher accuracy. However, we would also expect our runtime to greatly increase. Utilizing more computational power in the form of GPUs from Google Cloud can, therefore, greatly enhance the potential of the project and machine learning algorithm.

### 2. Incorporating Complex Preprocessing and Neural Network Architectures

In addition to expanding the neural network's architecture, another extension could involve using various preprocessing methods and more complex neural network training strategies. One extension could be using a convolutional or recurrent neural network in the training process, which might be able to recognize much more subtle patterns.<sup>2</sup>

### 3. Accounting for Speed and Accelerations

In reality, vehicles do not always travel with constant velocities, and not all vehicles travel at the same speeds. At traffic lights and intersections, vehicles tend to slow down in the case of a yellow or red light. This non-constant velocity may affect induction readings in an ILD, which is why an extension of this project may include a classification task involving non-constant velocities.<sup>5</sup>

Vehicles moving at much faster or much slower speeds may also induce a greater or lesser change in induction. This is something that could be considered as an extension to the results of our project, but would require many more simulations.

### 4. Using More Complex Models in Ansys HFSS

In our simulations, we modeled vehicles as stainless steel blocks. Modeling vehicles as much more complex objects, which could include adding wheels, may result in different induction readings. This

would result in a much longer simulation, as the computations associated with more detailed models would be more complex. However, the results obtained from more complex simulations may be more reflective of real-world scenarios and may show different patterns in the data.

### **5. Using a higher frequency and in-depth simulation in Ansys HFSS (more steps)**

In the simulations used in this project, we used a frequency sweep from 1 kHz to 20 kHz. Electromagnetic waves that propagate at these frequencies are classified as low-frequency waves. Using higher-frequency waves that propagate at frequencies that are classified as high-frequency waves can induce a current in the loop that is more noticeable to the neural network, allowing classification to be easier.

### Vehicle Classification using ML Algorithm

#### Importing Libraries

```
[ ] import pandas as pd
import tensorflow as tf
import keras
from keras import layers
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
```

#### Preprocessing

```
[ ] from google.colab import drive
drive.mount('/gdrive')

Drive already mounted at /gdrive; to attempt to forcibly remount, call drive.mount("/gdrive", force_remount=True).
```

```
[ ] ls
```

```
'Vehicle Classification.ipynb'  VehicleData3.csv
VehicleData2.csv                  vehicleData.csv
```

```
[ ] df=pd.read_csv("VehicleData3.csv")
```

```
[ ] df.head()
```

	EVENT	L1	L2	L3	L4	L5	L6	L7	LABEL
0	NaN	-5.420000e-07	-5.510000e-07	-6.480000e-07	-6.960000e-07	-8.050000e-07	-7.010000e-07	-5.550000e-07	motorcycle
1	NaN	-3.140000e-07	-6.980000e-07	-8.630000e-07	-8.470000e-07	-8.470000e-07	-6.380000e-07	-3.780000e-07	motorcycle
2	NaN	-5.100000e-07	-6.050000e-07	-7.200000e-07	-8.910000e-07	-8.860000e-07	-7.400000e-07	-3.960000e-07	motorcycle
3	NaN	-3.160000e-07	-6.170000e-07	-6.940000e-07	-8.270000e-07	-8.590000e-07	-9.080000e-07	-4.360000e-07	motorcycle
4	NaN	-6.240000e-07	-6.280000e-07	-6.550000e-07	-7.500000e-07	-7.570000e-07	-6.260000e-07	-5.480000e-07	motorcycle

```
[ ] X = df[["L1","L2","L3","L4","L5","L6","L7"]]
```

```
[ ] X=X*10E7
```

```
[ ] y = df["LABEL"]
```

```
[ ] y = pd.get_dummies(y)
```

```
[ ] from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15, random_state=40)
```

```
[ ] X_train = np.asarray(X_train).astype(np.float32)
y_train = np.asarray(y_train).astype(np.float32)
X_test = np.asarray(X_test).astype(np.float32)
y_test = np.asarray(y_test).astype(np.float32)
```

#### Training Neural Network Model using Tensorflow

##### Defining Neural Network Architecture in Tensorflow

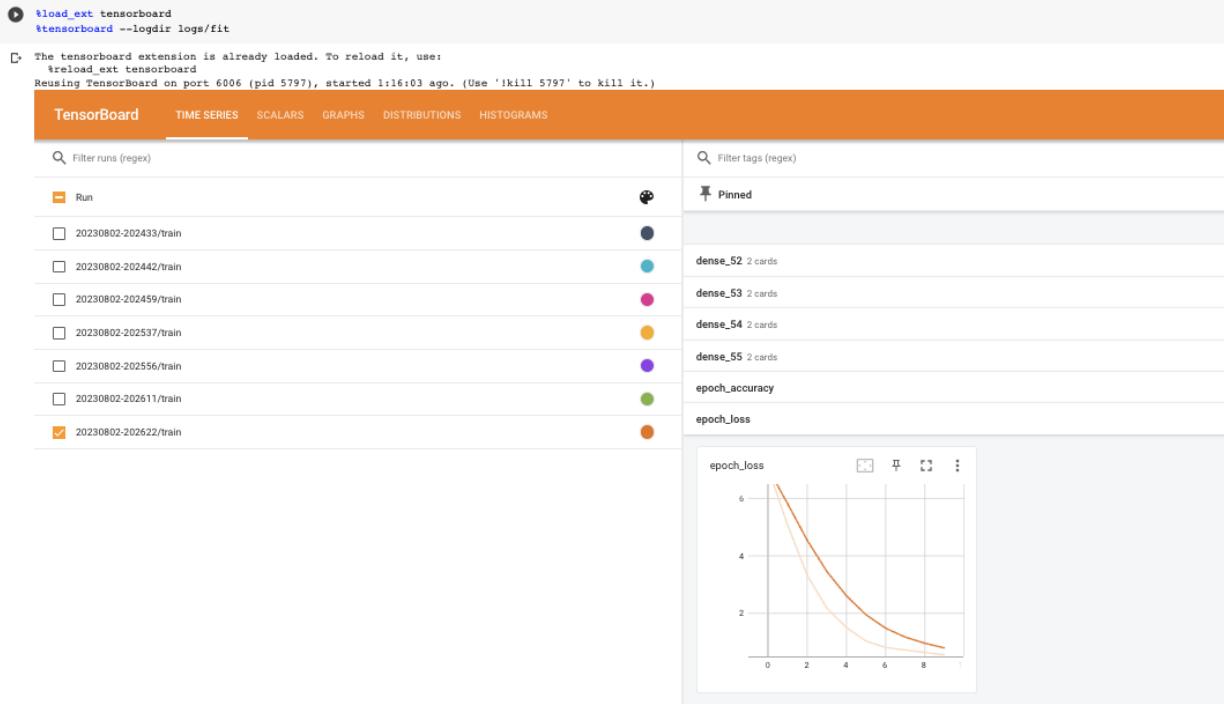
```
[ ] model = keras.Sequential()
model.add(layers.Dense(7, activation="relu"))
model.add(layers.Dense(64, activation="relu"))
model.add(layers.Dense(64, activation="relu"))
model.add(layers.Dense(3,activation="softmax"))

[ ] model.compile(loss = "categorical_crossentropy", optimizer = "adam",metrics=[ "accuracy"])

[ ] import datetime
log_dir = "logs/fit/" + datetime.datetime.now().strftime("%Y%m%d-%H%M%S")
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir=log_dir, histogram_freq=1)

[ ] model.fit(x=X_train,y=y_train,epochs=10, callbacks=[tensorboard_callback])

Epoch 1/10
1/1 [=====] - 1s 1s/step - loss: 7.0512 - accuracy: 0.3600
Epoch 2/10
1/1 [=====] - 0s 83ms/step - loss: 5.0756 - accuracy: 0.3600
Epoch 3/10
1/1 [=====] - 0s 95ms/step - loss: 3.3166 - accuracy: 0.4000
Epoch 4/10
1/1 [=====] - 0s 79ms/step - loss: 2.1737 - accuracy: 0.5600
Epoch 5/10
1/1 [=====] - 0s 76ms/step - loss: 1.5009 - accuracy: 0.6400
Epoch 6/10
1/1 [=====] - 0s 80ms/step - loss: 1.0274 - accuracy: 0.6000
Epoch 7/10
1/1 [=====] - 0s 94ms/step - loss: 0.8081 - accuracy: 0.7200
Epoch 8/10
1/1 [=====] - 0s 85ms/step - loss: 0.7205 - accuracy: 0.8400
Epoch 9/10
1/1 [=====] - 0s 82ms/step - loss: 0.6386 - accuracy: 0.8400
Epoch 10/10
1/1 [=====] - 0s 82ms/step - loss: 0.5491 - accuracy: 0.8400
<keras.callbacks.History at 0x7d4f48352a10>
```



```
[ ] model.predict(X_test)
[ ] [=====] - 0s 117ms/step
array([1.2634964e-01, 8.5844868e-01, 1.5201673e-02],
      [3.1460963e-10, 3.8371984e-02, 9.6162802e-01],
      [2.0937499e-10, 2.3984385e-01, 7.6615615e-01],
      [3.5597030e-09, 4.0323070e-09, 9.9995971e-01],
      [1.1669932e-10, 9.5763795e-02, 9.0423620e-01]), dtype=float32)

[ ] predictions=np.argmax(model.predict(X_train),axis=1)
[ ] [=====] - 0s 27ms/step

[ ] predictions
array([2, 1, 0, 1, 1, 2, 0, 2, 1, 1, 0, 1, 2, 1, 0, 2, 1, 2, 1, 1, 0,
       0, 1, 0])

[ ] np.argmax(X_train,axis=1)
array([6, 6, 0, 6, 6, 0, 6, 0, 4, 0, 6, 0, 0, 3, 6, 0, 0, 4, 6, 0,
       6, 0, 2])

[ ] from sklearn.metrics import classification_report

[ ] print(classification_report(np.argmax(y_train, axis=1), predictions))

          precision    recall  f1-score   support
          0       0.88      0.78      0.82       9
          1       0.82      0.90      0.86      10
          2       1.00      1.00      1.00       6

     accuracy                           0.88      25
   macro avg       0.90      0.89      0.89      25
weighted avg       0.88      0.88      0.88      25

[ ] predictions
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])

[ ] model.summary()
Model: "sequential_13"

```

Layer (type)	Output Shape	Param #
dense_52 (Dense)	(None, 7)	56
dense_53 (Dense)	(None, 64)	512
dense_54 (Dense)	(None, 64)	4160
dense_55 (Dense)	(None, 3)	195

```
Total params: 4,923
Trainable params: 4,923
Non-trainable params: 0
```

## VI. References

1. Electrical4u. (2012, February 24). *Faraday's Laws of Electromagnetic Induction: First & Second Law* | Electrical4U. Electrical4U. <https://www.electrical4u.com/faraday-law-of-electromagnetic-induction/>
2. Goodfellow, I. (2012). *Deep Learning*. Deeplearningbook.org. <https://www.deeplearningbook.org/contents/mlp.html>
3. IBM. (2023). *What is Gradient Descent?* | IBM. Wwww.ibm.com. <https://www.ibm.com/topics/gradient-descent>
4. Illinois Institute of Technology. (2015). *Experiment 11: Faraday's Law of Induction*. [https://www.iit.edu/sites/default/files/2019-11/faraday\\_s\\_law\\_mod.pdf](https://www.iit.edu/sites/default/files/2019-11/faraday_s_law_mod.pdf)
5. Kerner, B. S. (2004). Three-phase traffic theory and highway capacity. *Physica A: Statistical Mechanics and Its Applications*, 333(0378-4371), 379–440. <https://doi.org/10.1016/j.physa.2003.10.017>
6. Kingma, D. P., & Ba, J. (2014, December 22). *Adam: A Method for Stochastic Optimization*. ArXiv.org. <https://arxiv.org/abs/1412.6980>
7. LibreTexts. (2019, February 23). *6.8A: Electrical Conductivity and Resistivity*. Chemistry LibreTexts. [https://chem.libretexts.org/Bookshelves/Inorganic\\_Chemistry/Map%3A\\_Inorganic\\_Chemistry\\_\(Husecroft\)/06%3A\\_Structures\\_and\\_energetics\\_of\\_metallic\\_and\\_ionic\\_solids/6.08%3A\\_Bonding\\_in\\_Metals\\_and\\_Semiconductors/6.8A%3A\\_Electrical\\_Conductivity\\_and\\_Resistivity](https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Map%3A_Inorganic_Chemistry_(Husecroft)/06%3A_Structures_and_energetics_of_metallic_and_ionic_solids/6.08%3A_Bonding_in_Metals_and_Semiconductors/6.8A%3A_Electrical_Conductivity_and_Resistivity)
8. M. Pilar Velasco, Jose Luis Vazquez-Poletti, & Vázquez, L. (2022). About the Simulations of Maxwell Equations: Some Applications. *Springer Link*, 36, 45–54. [https://doi.org/10.1007/978-3-031-06632-0\\_3](https://doi.org/10.1007/978-3-031-06632-0_3)
9. Moreno, F. A. (2021, November 30). *Sparse Categorical Cross-Entropy vs Categorical Cross-Entropy*. Medium. <https://fmorenovr.medium.com/sparse-categorical-cross-entropy-vs-categorical-cross-entropy-ea01d0392d28>
10. Tensorflow. (2023). *Guide | TensorFlow Core*. TensorFlow. <https://www.tensorflow.org/guide>

11. U.S. Department of Transportation. (2006). *Traffic Detector Handbook: Third Edition-Volume I.*  
<https://www.fhwa.dot.gov/publications/research/operations/its/06108/06108.pdf>
12. Wikipedia. (2023a, January 30). *Three-phase traffic theory*. Wikipedia.  
[https://en.wikipedia.org/wiki/Three-phase\\_traffic\\_theory](https://en.wikipedia.org/wiki/Three-phase_traffic_theory)
13. Wikipedia. (2023b, May 9). *Electromagnetic induction*. Wikipedia.  
[https://en.wikipedia.org/wiki/Electromagnetic\\_induction#Applications](https://en.wikipedia.org/wiki/Electromagnetic_induction#Applications)
14. Wood, T. (2019, May 17). *Softmax Layer*. DeepAI. <https://deeppai.org/machine-learning-glossary-and-terms/softmax-layer>
15. Yuan, S. (2014). Identification and optimization of traffic bottleneck with signal timing. *Journal of Traffic and Transportation Engineering (English Edition)*, 1(5), 353–361.  
[https://doi.org/10.1016/S2095-7564\(15\)30281-6](https://doi.org/10.1016/S2095-7564(15)30281-6)

# Metal Detection Techniques Using Mutual and Self Inductance

Megan Grolemund, Jessica Hou, Trinity Kong

## Abstract

Metal detectors have been used for years with different designs and detection methods. Defense against unwanted metallic objects and treasure hunting are among the most common uses for metal detectors. This paper will explore the methods of mutual inductance and self-inductance with respect to metal detection, along with the theory supporting the methods. Hans Christian Oersted and Michael Faraday's respective experiments described the basis of electromagnetic induction. The process of building the metal detectors began with designing block diagrams of each part of the mutual inductance circuit and self-inductance circuit. Then, an online circuit simulator called Multisim was used to model the block diagrams, ensuring correct elements and connections. With functioning simulations, the building of physical breadboards began. The Multisim simulation of the breadboards functioned as expected and provided good data to implement the design onto the physical breadboards, however, the physical breadboards did not perform as anticipated. Assuming the connections were made correctly, the main issues concerned the coil and the circuit board elements. Future works could explore variations in the number of turns, diameter, and gauge of wire along with sensitivity of our circuit boards.

## I. Introduction

### A. Usages of Metal Detectors

Metal detection has been used for centuries in ways such as searching for landmines, metal weapons, and treasure. The first known metal detector was used in China about 200 years ago. The metal detector was a metal door, potentially magnetite, constructed to prevent people from carrying metal objects around the emperor, as they could potentially be dangerous weapons. The first electrical metal detector was designed by Alexander Graham Bell in 1881. The metal detector was used to locate a bullet shot into President James Garfield. During World War II, modern electrical metal detectors advanced rapidly in order to detect buried landmines. The 1970s brought about low-cost yet sophisticated handheld metal detectors due to developments in integrated circuits. The 1980s saw growth in the concept of metal discrimination, allowing treasure hunters to differentiate clutter objects from valuable objects.

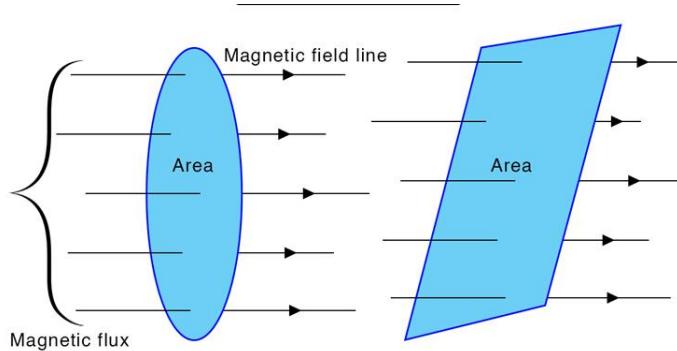
### B. Electromagnetic Induction

Metal detectors detect metal using electromagnetic induction. The concept of electromagnetic induction is derived from the knowledge of electrical currents and magnetic fields. Electric current refers to the net amount of charge that passes through a conductive material in a given unit of time.<sup>2</sup> The two general forms of currents are alternating current (AC) and direct current (DC). Alternating current refers to a current in which the direction alternates periodically, creating an oscillating wave. Conversely, the direction of a direct current remains constant, producing a straight line of current rather than an oscillating wave.<sup>3</sup> Magnetic fields are the region of space around a magnetic object or current-bearing material that moving charges interact with. These two concepts exhibit a symmetric relationship: flowing current produces a magnetic

field, and a change in the magnetic field induces a current.<sup>4</sup> This relationship is the basis for metal detection using self and mutual inductance.

The magnitude of magnetic fields can be described by magnetic flux, the total magnetic field density in a given area, shown by this equation:

$$\text{Flux} = \Phi = BA\cos\theta$$

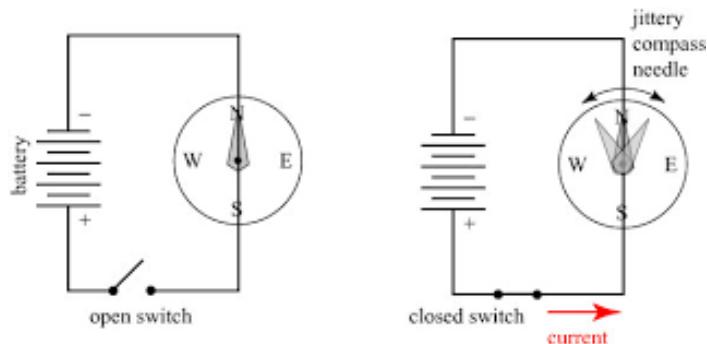


**Figure 1: Magnetic Flux<sup>5</sup>**

Figure 1 provides a visual representation of magnetic flux in relationship to the magnetic field lines. The greater the change in the magnetic field, the greater the strength of the magnetic field's induced current. The magnitude of magnetic flux depends on the strength of the magnetic field *B*, the area in which the magnet passes *A*, and the angle of the field lines with respect to the object  $\cos\theta$ . Any decrease in the magnetic field, area, or rearrangement of the perpendicular magnetic field lines to the object will decrease the induced current.<sup>5</sup>

## 1. Theory

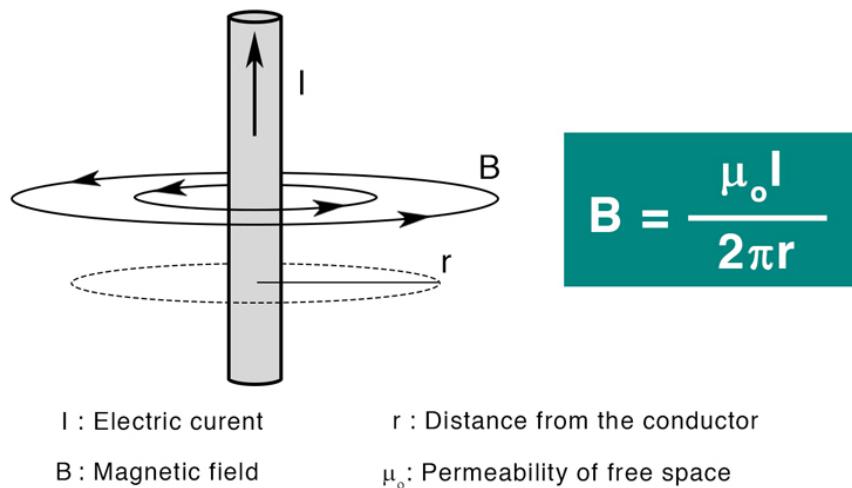
Before physicists realized that magnetic fields had the ability to induce current, Hans Christian Oersted discovered in 1820 that all current flowing through a wire generates a magnetic field.



**Figure 2: Oersted's Experiment<sup>13</sup>**

In Oersted's experiment shown in figure 2, he set up a current-bearing wire next to a compass. Compasses point in the direction of magnetic field lines, usually those of Earth's magnetic field. This changes when there is an additional stronger field present, such as ones from a nearby magnet.

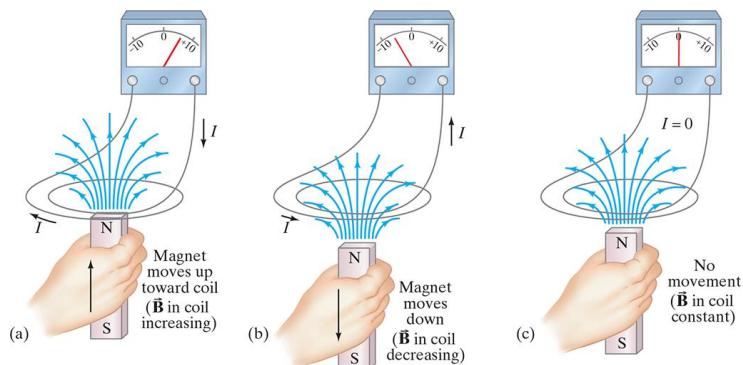
Once Oersted introduced a current in the wire, he expected the compass to continue to point north in the direction of Earth's magnetic field lines; to his surprise, the needle deflected away from Earth's field lines. When he switched the current off, the compass returned back to its original position. This led Oersted to the conclusion that electrical currents produce a magnetic field; this explains the magnetic field that the compass needle was pointing towards.<sup>6</sup>



**Figure 3: Visualization of Magnetic Field<sup>7</sup>**

Figure 3 gives a visual representation of Oersted's findings. The magnetic field flows perpendicular to the conductor and in a circular motion. The further  $r$  distance away from the conductor, the less strong the magnetic field will be.

Nearly a decade later, in 1831, Michael Faraday conducted further research on the relationship between electric current and magnetic field, concluding that magnetic fields can produce or induce currents.



**Figure 4: Michael Faraday's Experiment<sup>8</sup>**

Faraday's experiment consisted of a magnet, coil of wire, and galvanometer. Faraday began by moving a bar magnet through the wire coil connected to a galvanometer which measured electric current as shown in figure 4. He read through the galvanometer that the current increased or decreased from zero every time the bar magnet was inserted or removed. When the magnet was stationary in the coil, no current was induced. This can be explained by a change in magnetic flux.

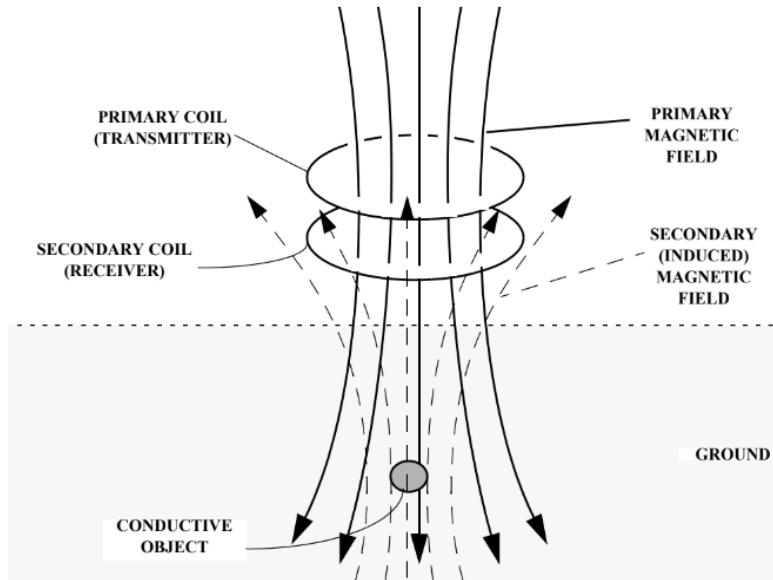
As the metal object moves further away from the conductor, the magnetic field lines weaken, and thus there is a decrease or change in magnetic flux. This change is what induces a current into the wire.

$$\varepsilon = -N \frac{\Delta\Phi}{\Delta t}$$

The above equation summarizes the conclusions of this experiment and explains Faraday's law of induction which states that a change in magnetic flux induces a current into a conductor. He additionally found that an increase in  $N$  number of turns increases the induced EMF. This showed a proportional relationship—electric currents generate magnetic fields, and magnetic fields induce current.<sup>9</sup>

### C. How Metal Detectors Work

Metal detectors are composed of a search head containing one or more coils that carry electric current and generate a magnetic field. The concept of metal detection relies heavily on Faraday's law of induction and is centered on the idea of analyzing a return signal from a transmitted magnetic field. Metal detection may use two types of induction: self and mutual.

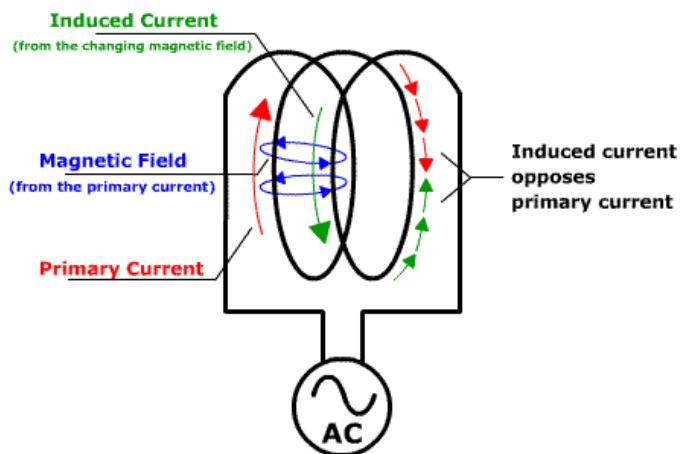


**Figure 5: Metal Detection Search Heads<sup>10</sup>**

For mutual inductance, the primary coil's magnetic field reacts with the properties of the conductive object and generates a secondary magnetic field. The secondary coil receives that magnetic field and gets its own induced current. Figure 5 illustrates the process and relationship between the magnetic field lines of the primary and secondary coil. The magnitude of the secondary magnetic field depends on the conductive

object's distance and orientation from the search head, its shape, size, conductivity, and permeability, and the presence of any background signal.

Self-induction consists of one coil with an alternating current. According to Oersted's experiment, this current-bearing coil will have a magnetic field. When a metallic object is brought in the presence of this magnetic field, it will experience a change in magnetic flux which will induce a current in the object. This current will produce a magnetic field which interacts with the primary current's magnetic field and induce a current flowing in opposition to the primary current shown by the green arrow in Figure 6.

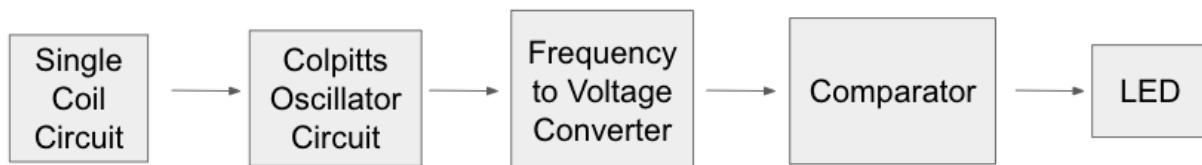


**Figure 6: Self-Inductance<sup>11</sup>**

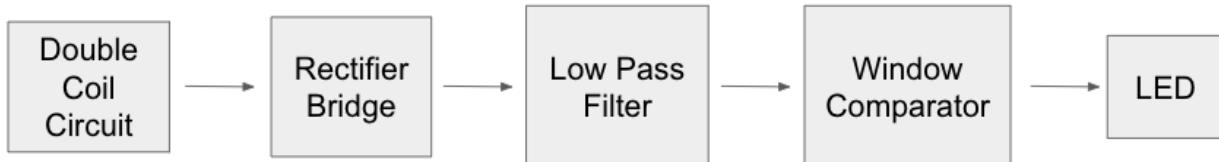
For example, if the current in a wire changes, the magnetic flux generates an induced current flowing in the opposite direction of the primary current. This phenomenon works only for alternating currents, which periodically reverse direction. The induced current in both cases is picked up by a receiver and then transformed into an output source known as voltage, electrical potential energy per unit charge, which travels down the circuit and gets converted into audio or light signals through a speaker or LED.<sup>12</sup>

## II. Block Diagrams and Final Design Elements

Diagrams of potential circuit setups to achieve metal detection were created for both self-inductance (single coil) and mutual inductance (double coil) cases.



**Figure 6. a. Self-Inductance Initial Design**

**Figure 6. b. Self-Inductance Final Design****Figure 6. c. Mutual Inductance Initial Design****Figure 6. d. Mutual Inductance Final Design**

## 1. Single Coil Circuit

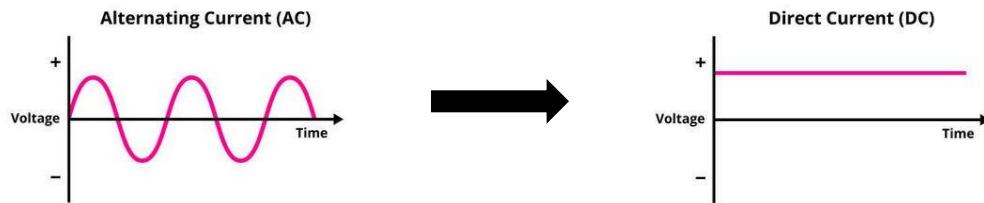
A single coil circuit is made of one coil of a conductive material, which is copper, in this specific experiment. Copper is used because of its large number of free electrons, which allow the metal to be a good conductor of electricity. The primary current generates a magnetic field that is able to induce a current in an introduced conductive material. The induced current creates its own magnetic field, which opposes the source current in the coil. The interaction between the induced and primary currents is the key to detecting a magnetic material near the metal detector.<sup>14</sup>

## 2. Double Coil Circuit

A double coil circuit works similarly to a single coil circuit, except it uses the source coil to send out a primary current, and the detection coil is used to detect the induced current of a nearby conductive material. The magnetic field of the primary current will induce a current and magnetic field in an introduced metallic object. The induced current of the metallic object then induces a current in the detection coil. To run a current through the circuit, the current must come from the smaller detection coil because it will have a smaller frequency than the larger source coil's current.<sup>15</sup>

### 3. Rectifier Bridge

In the mutual inductance detector, a rectifier bridge is used to translate an alternating current into a direct current:



**Figure 7: Alternating and Direct Current Graphs<sup>16</sup>**

The rectifier bridge is composed of four diodes, which are elements that allow current to flow easily from the negative end to the positive end but not from the positive end to the negative end. Two of the diodes are positioned to allow positive current to flow through, and the other set of two diodes are positioned to allow negative current to flow through. These two sets of diodes make the transition from alternating current to direct current possible because each set accounts for either the positive or negative part of the current.<sup>17</sup>

### 4. Low Pass Filter

Following the rectifier bridge in the mutual inductance detector, the low pass filter sets a cutoff frequency and only puts currents with the same or lesser frequency through the circuit. The cutoff frequency allows for only the detection coil's induced current to pass through the circuit, as the source coil's primary current would not meet the cutoff. This part of the circuit contains a capacitor, a circuit board element that holds charge, and a resistor, an element that limits the amount of current passing through.<sup>18</sup>

### 5. Non-inverting Comparator Circuit

The non-inverting comparator circuit is made of a potentiometer and an operational amplifier. A potentiometer consists of three resistors that divide the charge coming into the circuit part. This lessens the voltage going into the operational amplifier. The operational amplifier compares said input voltage with a reference voltage. If the input voltage is higher than the reference voltage, the output voltage will be positive. This circuit ensures that the correct input voltage will result in an amplified output voltage.<sup>19</sup>

### 6. 555 Oscillator Circuit

In the self-inductance design, the 555-oscillator circuit oscillates the incoming direct current in order for the actual detection of metal to work with the single coil circuit. The oscillation of current amplifies the current as well, making stronger magnetic fields and easier detection of magnetic materials. The circuit contains only a 555 timer, containing three resistors, two operational amplifiers, and one S-R latch.

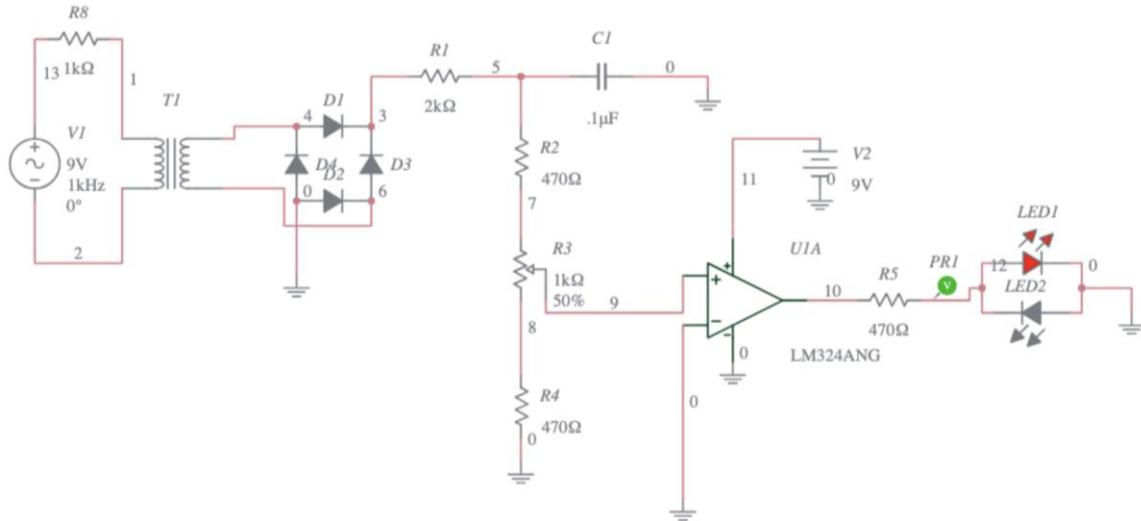
### 7. Led Light and Speaker/ Buzzer

The final parts of both the self-inductance and mutual inductance designs are either the led light or speaker/buzzer. These two parts show the results of their respective circuits, as in whether or not a metallic object was detected nearby. Both parts simply require the correct amount of voltage to flow through the circuit

towards them in order to work. The correct amount of voltage will be achieved when the nearby metallic object's magnetic field interacts with the current of the coils.

### III. Multisim Simulation

#### 1. Mutual Inductance Final Design



**Figure 8: Multisim Mutual Inductance Circuit**

The final Multisim design for mutual inductance is shown in Figure 8 and is based on Figure 6. d. The block diagram was turned into individual components all put together. Alternating current comes from the AC power source (V1) and goes to the primary coil. In this simulation, the metal is assumed to be already introduced, causing the secondary coil to have an induced current. That current then goes to the rectifier bridge (Figure 10. c.) and the AC is turned into DC. It then goes through the low pass filter (Figure 10. d.) and the non-inverting comparator (Figure 10. e.). If the voltage going in is greater than the set 9V (V2) then it will output a positive value causing the top LED to light up (LED 1).

#### 2. Mutual Inductance Troubleshooting and Testing

LED1 was turning on even when there was no voltage going through the circuit. An inductor-voltage source loop warning was fixed by adding a 1-ohm resistor to the beginning of the circuit. Different resistances were tested, and the potentiometer percentage was tested and changed. It was found that the potentiometer percentage is insignificant. The 5-terminal amplifier had to be changed into an LM324ANG amplifier. The LM324ANG works better with the range of voltage being worked with. A DC source and ground also had to be connected to the +VCC. Once the circuit was working, the voltage was changed to find the values of when the LED was on or off. The inductance value for the coils was based on the physical model's coils calculated by  $L = \frac{N^2 \mu_0 \mu_r A}{l}$ . The primary coil is 10mH and the secondary coil is 7.7mH.

### 3. Self-Inductance Final Design

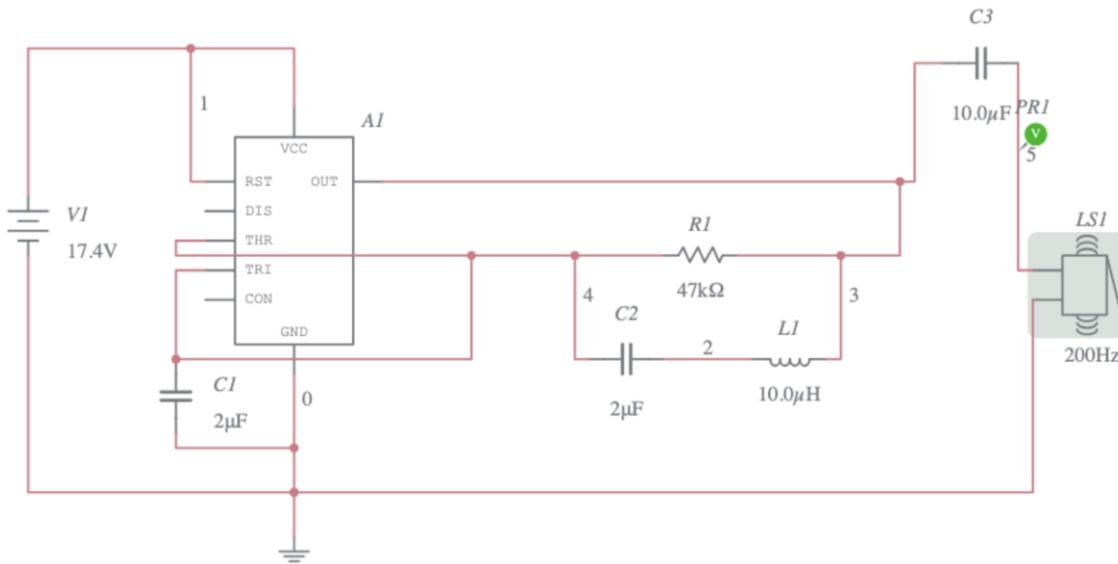


Figure 9: Multisim Self Inductance Circuit

The final Multisim design for self-inductance is shown in Figure 9 and is based on Figure 6. b. The block diagram was turned into individual components all put together. Direct current comes from the DC power source and goes through the oscillator circuit (Figure 10. f.) creating an oscillated wave. The oscillated wave then runs through the coil ( $L_1$ ) recreating an alternating current. In the circuit the presence of metal is assumed, so an induced current goes back through the coil and then to the buzzer (LS1).

### 4. Self-Inductance Troubleshooting and Testing

The buzzer had originally been an LED. The LED would blink and then stop blinking after running for a certain amount of time, depending on the input voltage. The voltage was graphed, and it was found that the LED caused a continuous decrease in voltage, causing it to turn off. On the other hand, the buzzer voltage plateaued and stayed at a certain voltage allowing it to stay on. The input voltage was then adjusted to find the values to make the circuit work. In the process, the inductance value of the coil and the voltage were the only variables that changed. These variables can be easily changed on the physical board. The inductance value was tested for the calculated inductance of the physical model, however, the final design was created using the theoretical value, and the physical model was changed to match.

### A. Breadboard Elements Used in Final Designs

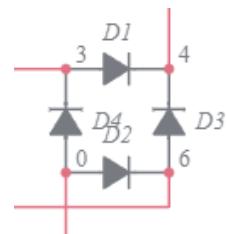
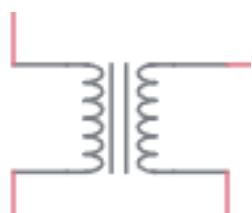
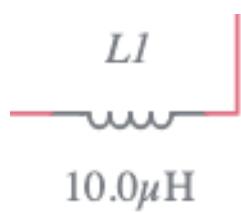


Figure 10. a. Single Coil Circuit

Figure 10. B. Double Coil Circuit

Figure 10. c. Rectifier Bridge



Figure 10. d. Low Pass Filter

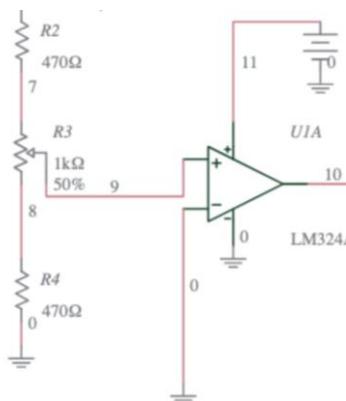


Figure 10. e. Non-Inverting Comparator Circuit

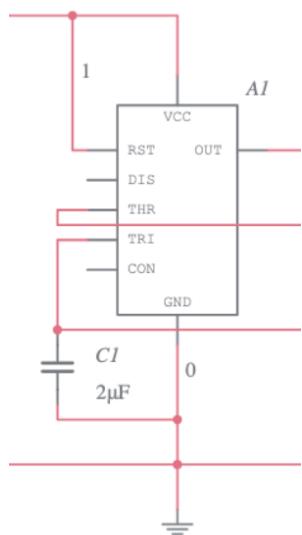


Figure 10. f. 555 Oscillator Circuit

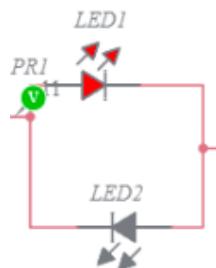


Figure 10. g. Led Light

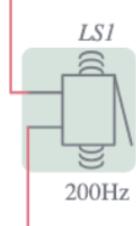


Figure 10. h. Speaker

## IV. Breadboard Design

### A. Mutual Inductance

#### 1. Initial Design

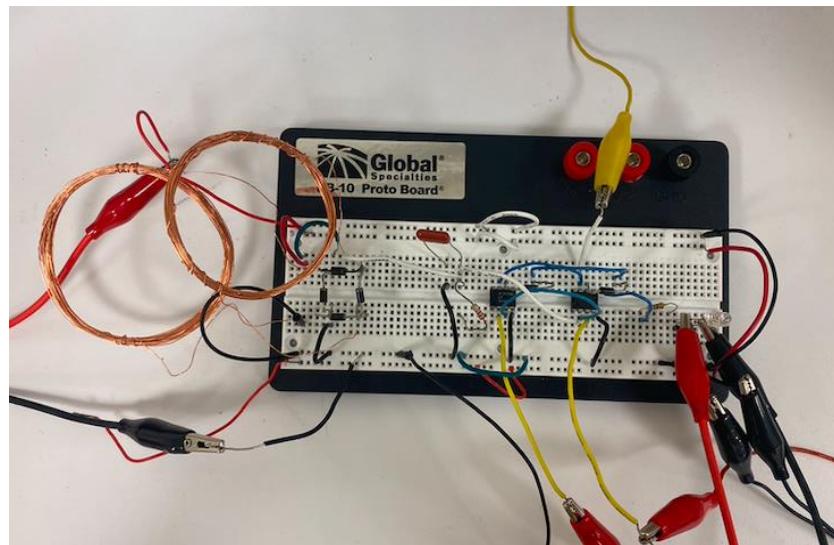


Figure 11: Mutual Inductance Breadboard Initial Design

For the mutual inductance detector, the initial design consisted of the double coil circuit, the rectifier bridge, the low pass filter, the window comparator, and the LED light.

#### 2. Final Design

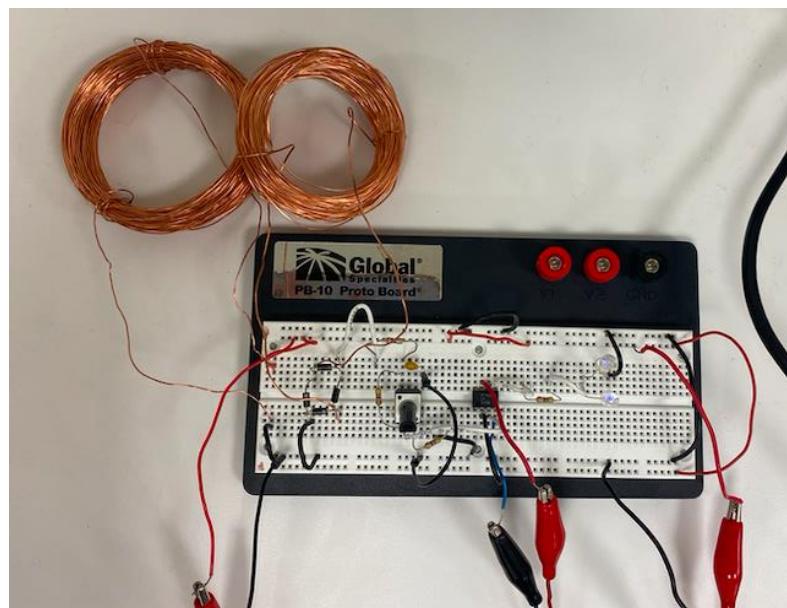
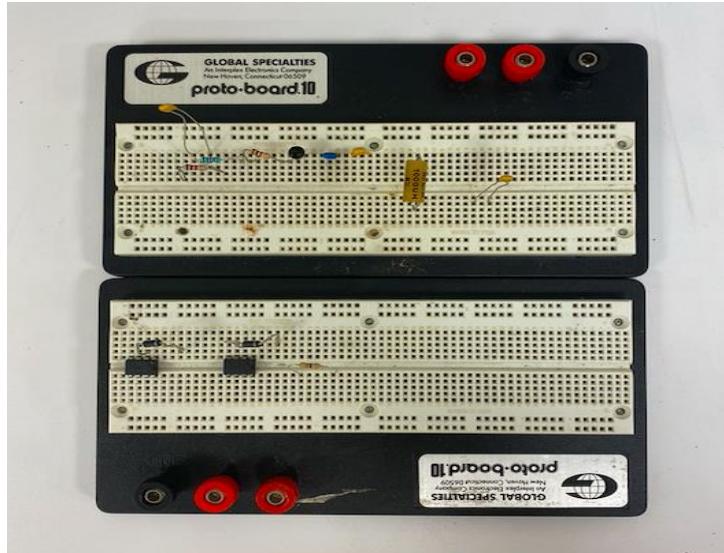


Figure 12: Mutual Inductance Breadboard Final Design

The next design for the mutual inductance breadboard consisted of two larger coils— both in the number of turns and in diameter, the rectifier bridge, the low pass filter, the non-inverting comparator, and the led light.

## B. Self-Inductance

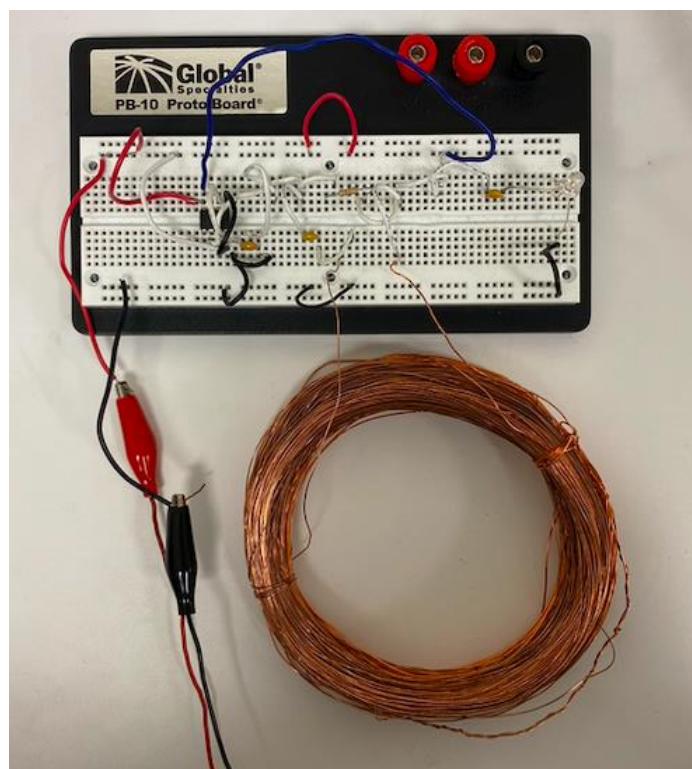
### 1. Initial Design



**Figure 13: Self-Inductance Breadboard Initial Design**

The first self-inductance detector design was never fully completed in its physical form as the simulator did not function properly, however, the Colpitts oscillator circuit (top) and the window comparator circuit (bottom) are displayed above.

## 2. Final Design



**Figure 14: Self-Inductance Final Design**

For the self-inductance breadboard, the final design contained the 555-oscillator circuit, the single coil circuit, and a speaker/ buzzer (although a LED light is depicted in the figure above).

## V. Results & Analysis

### 1. Analysis of Mutual Inductance Breadboard

The first mutual inductance breadboard design ran into problems with the window comparator circuit, and both breadboard designs lacked the sufficient number of turns in their respective coils and contained faulty circuitry. At the beginning of the breadboard construction, only 30 turns were used for both coils, which would not be sufficient for the needed amount of inductance. The 30 turns of the coil were increased to 200 turns of the coil, which was further increased to 400 turns in the second breadboard. The number of turns needed for the necessary inductance value was calculated by the following equation:

$$L = \frac{N^2 \mu_0 \mu_r A}{l}$$

L is the inductance value in Henries, l is the length of coil in feet,  $\mu_0$  is the constant for the inductance value of air,  $\mu_r$  is the constant for the inductance value of copper, N is the number of turns of wire, and A is the area inside of the coil.

In addition to the incorrect number of turns, the coils in the circuit were too small in diameter. The source coil began with a six-centimeter diameter, and the detection coil began with a five-centimeter diameter. The two diameters were then increased to ten and nine centimeters in diameter, respectively. The more copper wire used, and therefore the more turns and larger diameter of the coils, the more effective the magnetic field produced would be. Each coil turn generates its own magnetic field, so more turns would increase the coil's collective number of magnetic fields. Also, the strength of the magnetic field is directly proportional to the area of the coil, therefore, making the coils larger in diameter would increase the strength of the magnetic field, allowing it to detect metals over further distances. The window comparator circuit proved to be inefficient at its intended use. The two reference voltages and the range of voltages allowed to be put into the circuit were not specific enough, which may have caused issues with detecting the input voltage. This block of the design was changed to a non-inverting comparator circuit. Allowing only one reference voltage to be compared with the specific input voltage would be more efficient and accurate.

Finally, in both designs, the circuit's breadboard elements seemed faulty. Although they were tested with a multimeter to ensure proper functioning beforehand, certain elements did not function accurately when put into the breadboard and connected to other circuit components. A multimeter measures the voltage passing through a circuit component, whether it be a singular element, a part of a circuit, or an entire circuit. Using the multimeter again, the inserted breadboard components did not all display the correct input or output voltage. This could have been due to the breadboard itself having broken holes, as the breadboards used were not guaranteed to be fully functional. The nonfunctional components could have also contained faulty elements or faulty connections.

## 2. Analysis of Self Inductance Breadboard

Similar to the mutual inductance breadboards, the self-inductance breadboards lacked the correct number of turns and contained faulty elements. Using the equation shown below again, the first coil of 30 turns and a diameter of six centimeters was enlarged to a coil of 400 turns and a diameter of 10 centimeters.

$$L = \frac{N^2 \mu_0 \mu_r A}{l}$$

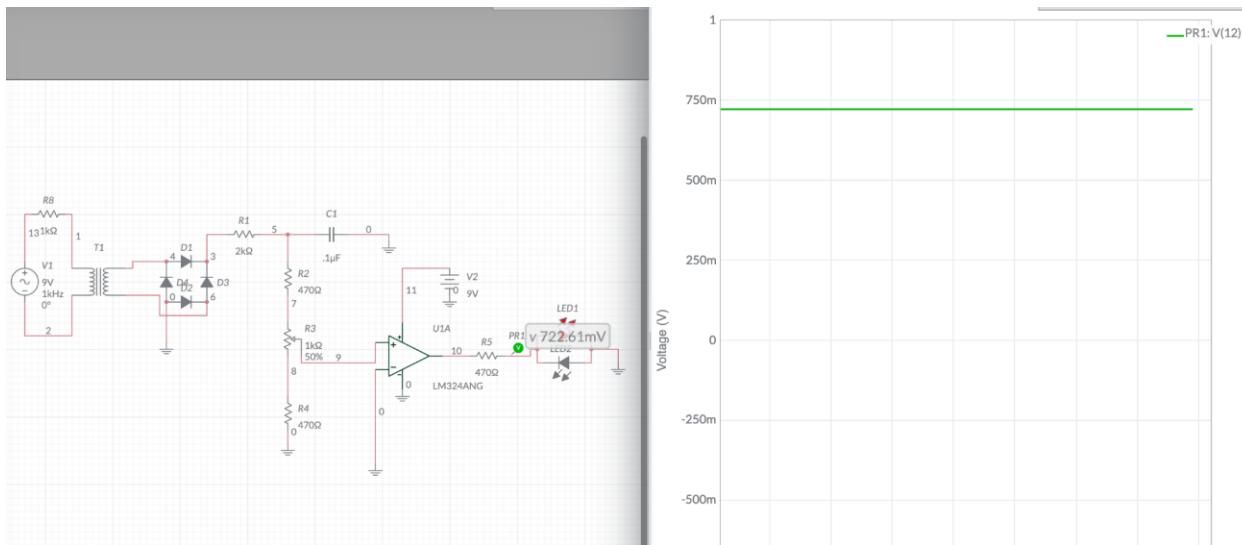
As for the faulty circuitry, the issue in the second breadboard could be narrowed down to one singular element, or the breadboard hole the element was inserted in. While testing the circuit connections with a multimeter, the very last capacitor was found to output a voltage of zero rather than the specific voltage that was input into the circuit board. Since every other element displayed the correct voltage in its input and output, it must have been the previously mentioned capacitor that was faulty.

However, the first design of the self-inductance detector also ran into more major issues. Presumably, because of the vast number of elements, and therefore connections, in the breadboard, there would be a high probability of incorrect connections and faulty elements. So, the design of the first breadboard was altered to condense three blocks of the design into one block: a 555-oscillator circuit.

## 3. Analysis of Mutual Inductance Multisim

The mutual inductance Multisim simulation worked properly. It was found that the LED was faintly on at 2.5 volts and was at full brightness at 5 volts. 633.25 mV was coming into the faint LED at 2.5 volts going into the circuit. 698.93 mV was coming into the fully bright LED at 5 volts going into the circuit. At 9 volts (theoretical input value), 722.61 mV went into the LED. It was also found that no maximum voltage value would break the circuit and be physically able to be put into the circuit. The graph (Figure 15) was a linear

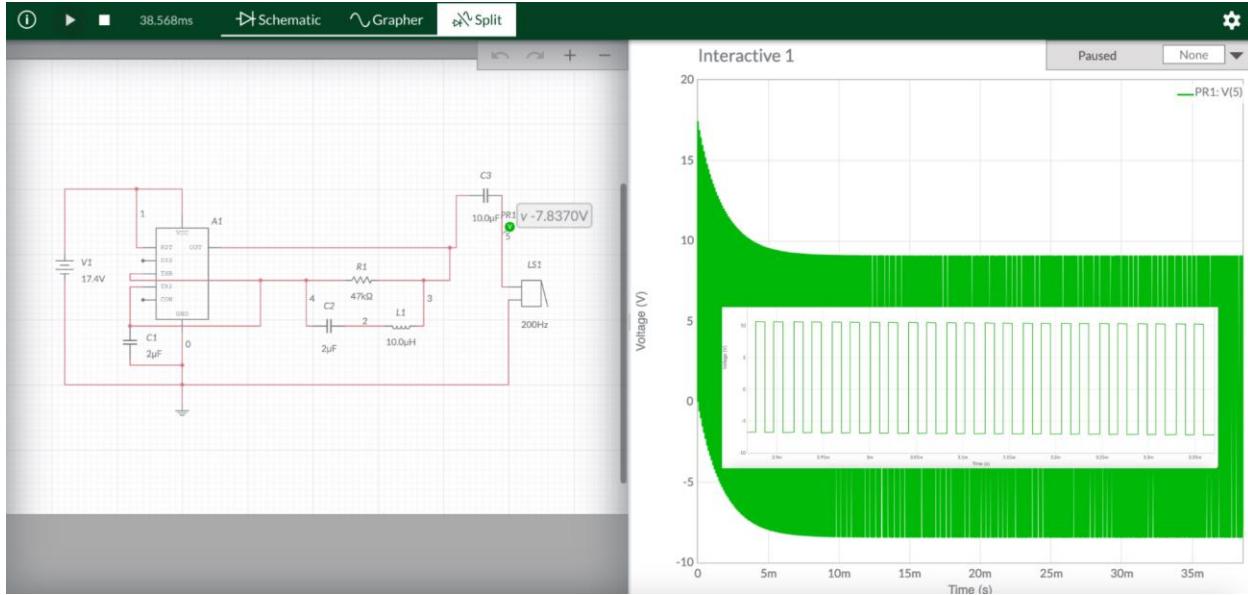
line because the rectifier bridge turned the AC at the beginning into DC. Only the top LED is on because the output value from the comparator is positive. If the bottom LED were on, that would mean that less than 9 volts would be coming from the other side of the circuit. If this were the case, it would most likely be a flaw in the rectifier bridge, allowing an alternating current to pass through.



**Figure 15: Multisim Mutual Induction Circuit and Graph**

#### 4. Analysis of Self Inductance Multisim

The self-inductance circuit simulation on Multisim worked properly. The LED continuous decrease can be demonstrated by comparing the decrease with the buzzer. For the LED,  $\Delta t = 131.634 \text{ ms}$ ,  $\Delta V = 51.57 \text{ mV}$ , and for the buzzer,  $\Delta t = 276.068 \text{ ms}$ ,  $\Delta V = 0.2 \text{ mV}$ . Therefore it is evident that there is a much greater decrease in the voltage when there is an LED in the circuit. The buzzer does not decrease over time at all, and the 0.2mV difference results from the inability to pinpoint the exact peak of each wave. A slight diagonal can be seen at the top of each wave in Figure 16. It was found that the buzzer needs 9 volts to sound. And 17.4 volts need to be put into the whole circuit for the buzzer to plateau at the needed 9 volts. This would mean that slightly less than 17.4 volts are needed for the physical circuit. When a metal is introduced to the coil, the voltage will increase slightly, putting the voltage at or above 17.4, making the buzzer sound.



**Figure 16: Multisim Self Inductance Circuit and Graph**

## VI. Future Considerations

Given the multitude of tests and breadboard designs, there were many components that would have been of interest to change and explore further. The following equation gives the value of inductance. Inductance can be increased by increasing the number of turns in the coil ( $N$ ), increasing the thickness, which would depend on the gauge of the wire and number of loops ( $l$ ), changing the material of the copper wire ( $\mu_r$ ), or increasing the radius of the loop.

$$L = \frac{N^2 \mu_0 \mu_r A}{l}$$

Inserting values from the self-inductance loop into the equation results in the following inductance value:

$$L = \frac{188^2 * 1 * 1.256624E - 6 * \pi(0.05)^2}{60.96}$$

$$L = 5.72 \mu H$$

This value of inductance is a possible reason for the lack of an output signal in our design. In the future, it is of interest to research the smallest value of inductance, which produces a return output. This can be done by incrementally increasing factors such as the number of turns or size of the coil and then measuring the strength of the induced current using a multimeter on the output pins. It would also be interesting to research the coil's sensitivity to a metal object from differing distances.

$$B = \frac{\mu_0 I}{2\pi r}$$

According to the magnetic field equation, objects  $r$  distance further from the center of the magnetic field receive an inversely lower magnetic field strength.

$$\phi = BA\cos\theta$$

Using the magnetic flux equation, a lower magnetic field results in a directly proportional lowered flux.

$$\varepsilon = -N \frac{\Delta\Phi}{\Delta t}$$

Combining the knowledge states that the farther the distance of the object from the center of a magnetic field, the lower the magnetic flux will be. Given that, further research can be done to test and compare the induced EMF values of objects in close proximity to the center of the magnetic field from distances that are further away. As long as the change in magnetic flux stays the same, the induced EMF should remain the same regardless of the distance from the center. This exploration could also be a way to measure the length of the magnetic field of an object where the end of the object's magnetic field is determined when the induced current ceases to be present.

## VII. Acknowledgments

We would like to thank Professor Chayanika Roy Chaudhuri, our team project advisor, and Léa Tonnerre, our team project teaching assistant, for guiding us through everything about our team project, from learning about the basics of inductance to troubleshooting the various issues we ran into along the way. Additionally, we want to thank Dr. Barry Luokkala, and the Pennsylvania Governor's School for the Sciences Campaign Inc. for the opportunity and the resources to participate in this project.

## VIII. References

1. Bruschini, Claudio, "A Multidisciplinary Analysis of Frequency Domain Metal Detectors for Humanitarian Demining" (2002). *Global CWD Repository*. 1295. <https://commons.lib.jmu.edu/cisr-global cwd/1295>
2. Urone, Paul Peter, and Roger Hinrichs. "Induced Emf and Magnetic Flux." *College Physics*, OpenStax, Houston, TX, 2022, <https://openstax.org/books/college-physics/pages/23-1-induced-emf-and-magnetic-flux>.
3. Earley, Elizabeth. "AC and DC Current." *MIT School of Engineering*, 17 Sept. 2013, [engineering.mit.edu/engage/ask-an-engineer/whats-the-difference-between-ac-and-dc/](https://engineering.mit.edu/engage/ask-an-engineer/whats-the-difference-between-ac-and-dc/).
4. "Electric & Magnetic Fields." *National Institute of Environmental Health Sciences*, [www.niehs.nih.gov/health/topics/agents/emf/index.cfm](http://www.niehs.nih.gov/health/topics/agents/emf/index.cfm).
5. "Magnetic Flux." *Science Facts*, <https://www.sciencefacts.net/magnetic-flux.html>. Accessed Sept. 2022.

6. Sack, Harald. "Hans Christian Ørsted Connecting Electricity and Magnetism." *SciHi*, 9 Mar. 2018, [scihi.org/hans-christian-orsted-electricity-magnetism/](https://scihi.org/hans-christian-orsted-electricity-magnetism/).
7. *Magnetic Field Equation*. <https://www.sciencefacts.net/magnetic-field.html>. Accessed 9 Mar. 2022.
8. *Michael Faraday's Experiment*. 4 Mar. 2022, <https://notes.mehvix.com/physics-7b/29/>.
9. Brock, Chip. "Lesson 11. Faraday's Experiments." *QS&BB*, [qstbb.pa.msu.edu/ed/lectures/L11\\_Faraday.2/](http://qstbb.pa.msu.edu/ed/lectures/L11_Faraday.2/).
10. *Mutual Inductance Diagram*. <https://www.texasgateway.org/resource/69-inductance>.
11. Davis, C. L. *Self Inductance Diagram*. [https://www.physics.louisville.edu/cldavis/phys299/notes/mag\\_selfind.html](https://www.physics.louisville.edu/cldavis/phys299/notes/mag_selfind.html).
12. Hughes, Scott. "Mutual and Self Inductance." Lecture 15. Massachusetts Institute of Technology Department of Physics, 7 Apr. 2005, Cambridge, Massachusetts.
13. Safarika, Eleftheria. *Hans Christian Oersted's Experiment*. <https://eimai.org/youth-shift-the-blog/2018/11/8/two-experiments-that-electrify-our-attention-and-magnetize-our-interest-oersteds-and-faradays-experiments>.
14. "Self-Inductance and Inductive Reactance." *Nondestructive Evaluation Physics : Electricity*, [www.nde-ed.org/Physics/Electricity/selfinductance.xhtml](https://www.nde-ed.org/Physics/Electricity/selfinductance.xhtml). Accessed 4 Aug. 2023.
15. "Nondestructive Evaluation Physics : Electricity." *Nondestructive Evaluation Physics : Electricity*, <https://www.nde-ed.org/Physics/Electricity/mutualinductance.xhtml>. Accessed 4 August 2023.
16. James, Luke. "ELECTRICITY BASICS What's the difference between AC and DC power?" *Power & Beyond*, 19 March 2020, <https://www.power-and-beyond.com/whats-the-difference-between-ac-and-dc-power-a-0c5c48e598b5e1266e6cebc5731227c2/>. Accessed 4 August 2023.
17. Last Minute Engineers. "The Full-Wave Bridge Rectifier." *Last Minute Engineers*, 15 May 2022, [lastminuteengineers.com/the-full-wave-bridge-rectifier/](https://lastminuteengineers.com/the-full-wave-bridge-rectifier/).
18. "Low-pass Filters | Filters | Electronics Textbook." *All About Circuits*, <https://www.allaboutcircuits.com/textbook/alternating-current/chpt-8/low-pass-filters/>. Accessed 4 August 2023.
19. "Non-Inverting Comparator | Analog-integrated-circuits || Electronics Tutorial." *Electronics-Tutorial.net*, <https://www.electronics-tutorial.net/analog-integrated-circuits/op-amp-comparators/non-inverting-comparator/>. Accessed 4 August 2023.

**PHYSICS**  
**TEAM PROJECTS**



# **Long Ago, in a Globular Cluster Far, Far Away: Discovering the Secrets of the Universe with the Messier 13 Globular Cluster**

Claire Thomas, Izabella Gurreonero, Keira Seidman, Lihini Ranaweera, Thomas Hasty

## **Abstract**

Globular clusters are a type of celestial object that can be used to provide insight into the history of the universe. Using a telescope housed at University of Pittsburgh's Allegheny Observatory, we imaged the globular cluster M13 to determine its age. By using images of M13 taken in different telescope filters, in conjunction with bias and flat calibration frames, we calculated the absolute magnitudes of 900+ stars in the cluster and used this information to create a Hertzsprung-Russell diagram (H-R diagram) for M13. With the data from the H-R diagram we were able to consult an isochrone chart to determine the age of M13. We also imaged various other celestial objects and used three gray-scale images of them taken with different filters to create three-color images and we describe the different structures of galaxies and various other astronomical phenomena that may have impacted our data. Using these techniques, we discover that the insights gained from studying globular clusters contribute to a deeper understanding of the universe.

## **I. Introduction**

### **A. What are Globular Clusters?**

#### **1. Introduction to Star Clusters**

The universe remains vast and mostly unexplored, with experts only having discovered a meager four percent of the apparent space. To compensate for our lack of knowledge, astronomers aim to use star clusters to chart the interaction between galaxies and make new discoveries about the nature of our universe. Within our research, there are two different types of star clusters which were discussed: open clusters and globular clusters

#### **2. Globular clusters vs. Open Clusters**

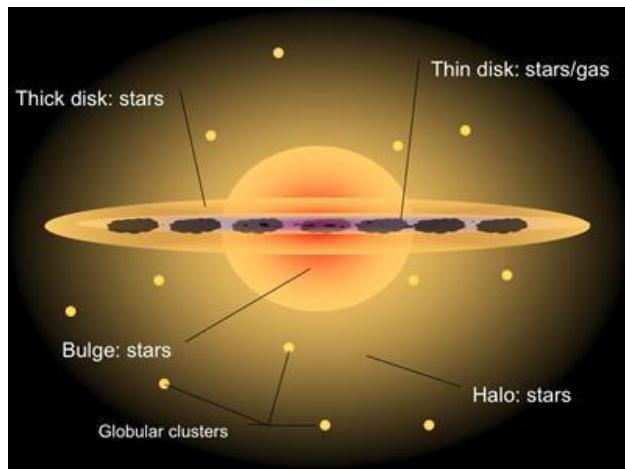
Open clusters are much smaller than globular clusters. These types of star clusters are scattered throughout the disk of our galaxy and presumably numerous others which have yet to be explored. Open clusters contain between a few dozen and a few thousand stars, which are all formed from the same initial cloud of gas and dust, and they have been observed to have a wide age range, likely being made up of younger or older stars. These star clusters have a density low enough that the individual stars within the open cluster can be observed with a telescope and, at times, even the unaided eye. This is what gives this genre of clusters their “open” appearance<sup>1</sup>. One will often find open clusters to be in the arms of spiral galaxies, with their stars being relatively young. In a steep contrast to globular clusters, open clusters cannot maintain a spherical shape as their structure is inherently unstable and are thus more irregularly proportioned. As such, these stars often disperse as the cluster rotates around the galaxy and are intercepted by the gravitational disruptions of passing cosmic objects.



**Figure 1: Open cluster Trumpler 14** NASA, ESA, and J. Maíz Apellániz (Institute of Astrophysics of Andalusia, Spain); Acknowledgment: N. Smith (University of Arizona)<sup>2</sup>.

Our research focused on finding the age of globular clusters and considering today's astronomical discoveries. Globular clusters are vastly different from open clusters, most glaringly due to the nature of their composition. Due to the high density of the stars at the center of these spherical clusters, they are hard to discern even with the most powerful telescopes. The large number of stars in a relatively small area leads globular clusters to appear as spherical, though the circumference of these clusters cannot be easily deduced due to the foreground and background cosmological bodies<sup>3</sup>.

The results of research led by countless astronomers all conclude that the stars within globular clusters are irrefutably old. Due to changes in luminosity and magnitude as a star ages, the eldest stars tend to be red, while the younger stars tend to be blue. The end of a star's lifespan can be especially violent, resulting in supernovas. Globular clusters are also found in all types of galaxies. Within the perimeter of the Milky Way, globular clusters are found within the *stellar halo*, a spherical population of stars. These stellar halo's surround most disk galaxies and the "cd" class of elliptical galaxies. An estimated one percent of a galaxy's stellar mass resides within this halo<sup>4</sup>. Observing halos in other galaxies is extremely difficult due to its low luminosity. Halo stars residing in the Milky Way are generally old, believed to be an estimated age of 12 gigayears.



**Figure 2: A typical construct of the spiral galaxy has a faint, extended stellar halo<sup>4</sup>.**

Understanding the age of globular clusters is a key parameter involved in understanding their relationship with galaxy formation and for testing model predictions about the universe<sup>5</sup>. When viewed with the naked eye, globular clusters appear to be faint smudges of light against the darkness of space. However, new telescopes, such as the James-Webb telescope, help resolve these clusters so that they are capable of being imaged and analyzed.



**Figure 3: Thousands of stars within globular cluster Terzan 4, part of the constellation Scorpius. The bright, blue stars shown in the center burn hotter and die quickly. The red stars on the outskirts of the cluster burn less hot and are much cooler.<sup>2</sup>**

### 3. Understanding globular clusters

Some globular clusters are found a very large distance away from the galactic center, which is the rotational center, or barycenter, of the Milky Way galaxy. Thus, globular clusters are now focused to be the key object for the study of distant parts of the Milky Way galaxy<sup>6</sup>.

Globular Clusters are fossils of the early development of the universe, though not much is known about their origin of creation. It is the gravitational attraction between these stars within globular clusters that allow them to become so long-lived, as they travel in a group together throughout different galaxies. According to the Harvard and Smithsonian Center for Astrophysics, these globular clusters are estimated to have formed around 10 billion years ago, which makes them some of the oldest stars in existence<sup>7</sup>.

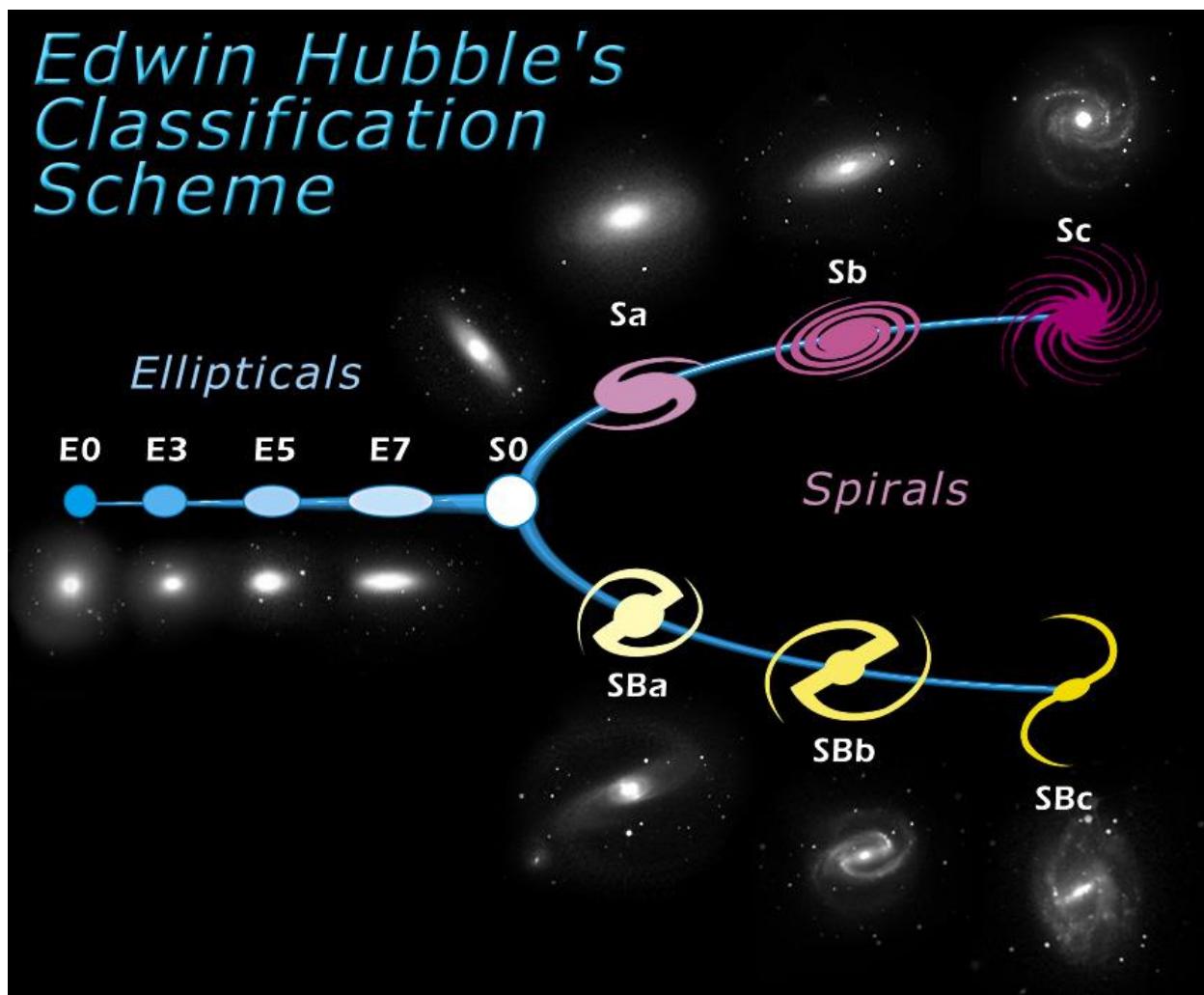
Since these stars within globular clusters are so old, they are “metal poor” meaning they lack heavy elements that did not exist in the early universe. These elements would be later created by supernovas. Globular clusters are lacking substantially in gas and dust, so the possibility for these clusters to produce new stars is highly unlikely.

One-way researchers can estimate the age of a globular cluster by its metallicity, better defined as the relative amounts of elements heavier than helium. The trend follows that younger clusters have a higher metallicity, while old clusters are “metal-poor”, as they were formed before supernovas provided the cosmos with heavier elements to absorb<sup>8</sup>. The stars contained in globular clusters are more condensed than any other place in the galaxy, which leads to the creation of binary systems, including those with compact objects: white dwarfs, neutron stars, and black holes.

## B. Structure of Galaxies

### 1. The Hubble Tuning Fork for Classification of Galaxies

Globular clusters are often an integral part of galaxies. The classification of galaxies based on their structure has evolved throughout the decades. Originating in 1926, the Hubble tuning fork for classification of galaxies has formed the backbone of galaxy classification in recent astronomy. The system categorizes most galaxies as being either elliptical or spiral. Edwin Hubble originally believed that the progression demonstrated the aging process of galaxies. Although this assumption turned out to be incorrect, some of its principles remained, specifically with the nomenclature associated with the system. The terms “early” and “late” are occasionally used to indicate the position of a galaxy on the tuning fork, with “early” indicating that the galaxy is towards the left, and “late” indicating the right. The system has remained a basis for other classification systems through its division between elliptical and spiral galaxies. It begins with a grouping of elliptical galaxies that progresses according to the ellipticity, then branches off into two separate groups: normal spirals and barred spirals<sup>9</sup>.



**Figure 4: A visualization of Edwin Hubble's Classification Scheme<sup>10</sup>.**

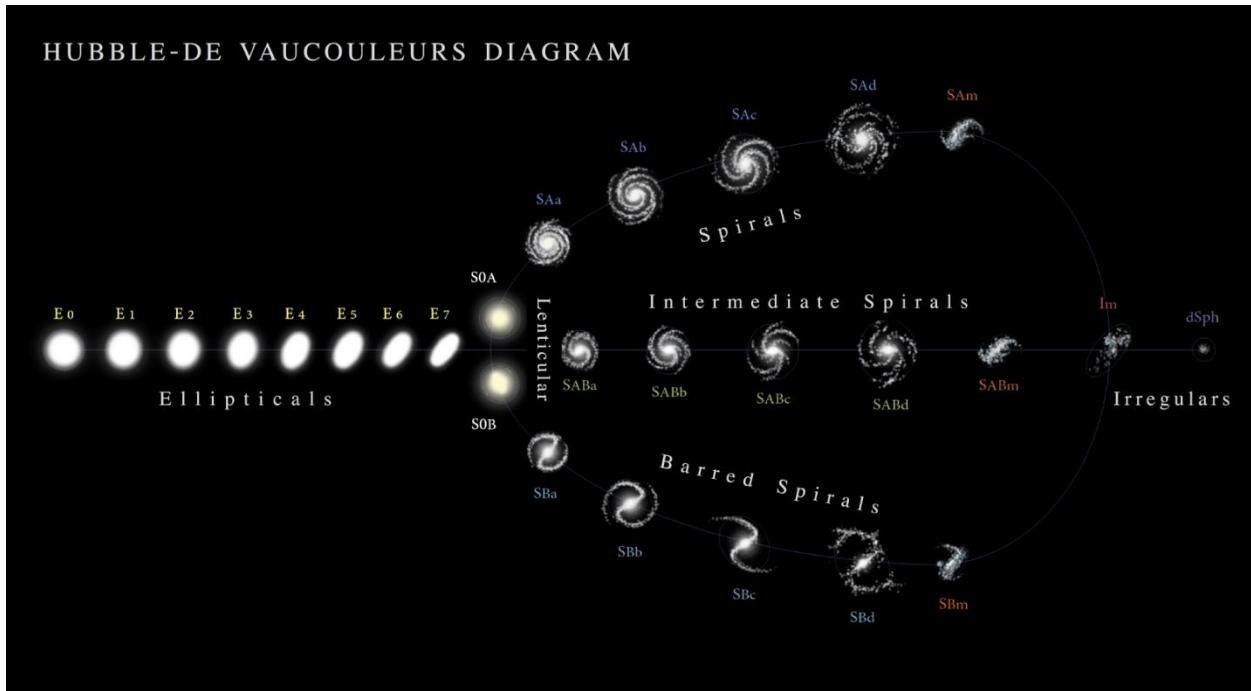
Spiral galaxies are labeled based on spiral compactness with letters "a-c" in the Hubble tuning fork, with "a" being the most compact. Spiral galaxies are also further classified into two groups: normal spirals and barred spirals. The centers of barred spirals appear to have a bar from which spirals originate, while normal spirals do not contain a bar. A bar is typically a ribbon of stars, gas, and dust that cut across the center of a spiral galaxy. The presence of bars is thought to indicate that a spiral galaxy has reached full maturity. Barred spirals comprise about % of the spiral group. Normal spirals are indicated with an "S" followed by their compactness level in the labeling system, and barred spirals are indicated with a "SB." Globular clusters are also an integral part of spiral galaxies and are specifically found in arms. These celestial bodies can help suggest the age of a spiral galaxy based on the age of the cluster itself<sup>11</sup>.

Elliptical galaxies, which reside on the left side of the tuning fork, are labeled based on an ellipticity from 0-7, with the "0" label indicating that the shape of the galaxy is a near-perfect circle. These galaxies are denoted with an "E" before its ellipticity value. These galaxies contain little gas or dust, and they have very little structure. Because there is not enough gas to form new stars, elliptical galaxies are typically populated with older stars that orbit in random directions. These galaxies are less common than spiral galaxies and are theorized to originate from collisions and mergers with spiral galaxies<sup>11</sup>.

Lenticular galaxies are another unique category of galaxies and are indicated with S0 as the transition zone between ellipticals and spirals. These are hybrids between elliptical and spiral galaxies. They contain the central bulge and disk common to spiral galaxies, yet have very little dust and star formation, which is characteristic of elliptical galaxies. It is theorized that lenticular galaxies could be older spirals whose arms have faded, or that they form from mergers between spirals<sup>12</sup>.

## 2. The de Vaucouleurs Classification System

Although the Hubble tuning fork classification system has been useful, it is considered to be an oversimplification, as it doesn't take into account various irregular galaxies. These include galaxies with odd shapes, small dwarf galaxies, and giant elliptical galaxies that reside in the centers of some clusters of galaxies. More modern galaxy classifications take into account the potential irregularities seen throughout the universe. One of the most widely used modern classifications is the de Vaucouleurs Classification system, which is an extension of the Hubble classification system. This system adopted new classifications based on Hubble classifications to denote a transition from well-developed spiral arms to more chaotic structures. This can be seen with the SA and SB structures in the de Vaucouleurs that correspond to Hubble types S and SB, as well as de Vaucouleurs' creation of the "d" and "m" subdivisions within the spiral group. These subgroups extend the way that spiral galaxy compactness is expressed and consider Magellanic Irregulars with the "m" subdivision. De Vaucouleurs also added SAB classifications for oval distortions, which provides a smoother transition between barred and unbarred spirals. Additionally, types "ab," "bc," "cd," and "dm" were adopted to denote smoother transitions from tightly wound to loosely wound spiral structures. Labels (r) and (s) were added to describe the presence and prominence of ring and/or spiral features. Smaller groups that take into account irregular galaxies were also added. Thus, this system broadens the scope of the Hubble tuning fork classification by adding several subgroups that take into account more irregular and unconventional forms of galaxies<sup>12</sup>.



**Figure 5:** A visualisation of Hubble-De Vaucouleurs Diagram<sup>13</sup>.

## C. Background on H-R Diagrams

### 1. Constructing an H-R Diagram

An H-R diagram is one of the most important data analysis tools available to an astronomer. Pioneered in the early 20th century by Ejnar Hertzsprung and Henry Norris Russel, an H-R diagram is a scatter plot that graphs the luminosity of a star against its temperature. Alternatively, it can plot the absolute V magnitude against the (B-V) color of each star<sup>14</sup>.

It is known that the hotter a star burns, the more energetic the photons it emits will be. This means that the light of a hotter star will be bluer than that of a relatively cooler star. By extension, when a hotter star is observed, its intensity will be greater in images taken with bluer filters. Similarly, cooler stars would have greater intensities in images with redder filters<sup>14</sup>.

Taking the difference in magnitude of a star between its appearance in a bluer filter versus a redder filter, yields a number that relatively ranks the temperature of a star. By general convention this is done with a blue filter and a Visual filter to get the (B-V) color. A visual filter is a type of green filter whose name is derived from the fact that green is the color that the human eye is most perceptive to<sup>14</sup>.

When calculating the (B-V) color value, it can be observed that hotter stars have smaller (B-V) color values and cooler stars have larger (B-V) color values. Magnitude is inversely related to the intensity of light, so a hotter star will have more blue light and a smaller B magnitude, making the (B-V) color value smaller. The opposite is true for cooler stars<sup>14</sup>.

The V magnitude can be used in place of luminosity precisely because it is the color that the human eye is most sensitive to. It gives a good metric of how bright the star would appear to the human eye, the basis of luminosity<sup>14</sup>.

## 2. Anatomy of an H-R Diagram

Once a collection of stars is plotted on an H-R diagram, a certain trend will become glaringly obvious. Figure 6 shows a general H-R diagram. This line that most stars fall on is called the main sequence. It shows that in general, as the temperature of a star increases, so will its luminosity. It is also observed that most stars in the main sequence differ in mass by no more than one order of magnitude from the sun, therefore the stars in the main sequence can be considered similar in mass<sup>14 15</sup>.

There are however some notable exceptions to the main sequence trend. Generally, in the top right of an H-R diagram there is a group of stars which appear to burn too bright for their temperatures. These stars are giants and supergiants, stars near the end of their life which have massively expanded. Additionally, in the bottom left of an H-R diagram are the dwarf stars. These are the leftover cores of stars that have already died. They appear too dim for their measured temperature. The luminosity of a star is proportional to the radius squared and the temperature to the 4th power and the H-R diagram is the visual representation of stellar evolution.

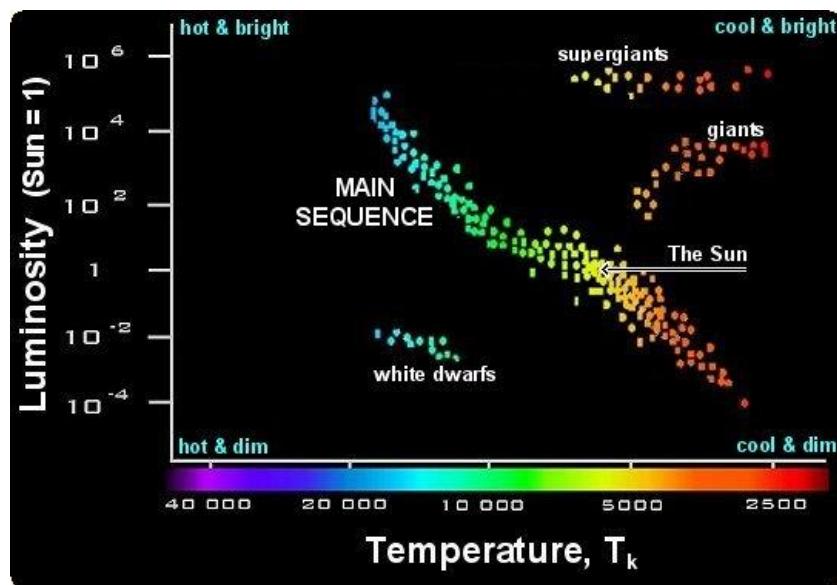


Figure 6: General H-R Diagram with Major Groups Labeled<sup>16</sup>.

## II. Telescope and Data Collection

### A. Allegheny Observatory and PlaneWave Telescope

The telescope used for our research, which can be seen in Figure 7, was a 24-inch PlaneWave telescope, which has a 4K X 4K CCD, a seven-element color wheel, and an f/6.5- and 3974-millimeter focal length. It is the fourth generation Keeler Memorial Telescope and was installed at the historic Allegheny Observatory

on August 25, 2022. The current observatory building was built in 1912 and is owned by the University of Pittsburgh.



**Figure 7: Fourth Generation Keeler Memorial PlaneWave Telescope at the Allegheny Observatory.**

As an institution, the observatory was created in 1859, and its first telescope was a 13-inch Fitz refraction telescope. The observatory later obtained a transit telescope, which allowed it to track the positions of the stars to make precise calculations about the time. This information was very valuable to many industries, especially the railroads, and became an important source of income for the observatory<sup>17</sup>.

The largest telescope in the observatory is the 30-inch Thaw refraction telescope. For many years, this telescope remained at the cutting edge of observational technology and was the primary instrument of the parallax program, which sought to measure the distances to various stars in the universe to provide some scale for the universe. Parallax involves imaging a star in six-month intervals to observe the star's shift in the sky from opposite sides of the Earth's orbital path. These shifts can be used in a trigonometric calculation to ascertain the star's distance from the Earth<sup>17</sup>.

## B. Data Collection

To use the PlaneWave telescope, we remotely connected to it via the University of Pittsburgh's secure VPN from Carnegie Mellon University's Doherty Hall. Once connected to the Observatory's computer system, we were able to operate both the telescope and observatory dome using various software programs, which are referenced in Appendices A – D.

Our observation was conducted over the course of three nights (7/10/23, 7/11/23, and 7/22/23) from around 9:00 pm – 1:00 am. These nights were chosen in particular because there was minimal cloud cover over Pittsburgh. After connecting to the telescope and opening all the required software, we began each night by taking bias frames and flat frames, which are two types of calibration frames that will be explained in a proceeding section. We also focused the telescope using the horizon line.

We then imaged five celestial objects as shown in Table 1. These included the primary focus of our research, the globular cluster M13; the star, Arcturus; the open cluster, M39; the nebula, NGC 6826; and the Veil Nebula. These images were taken in various filters (r', g', Blue, OIII, and H $\alpha$ ) which only allow certain wavelengths of light to be observed as shown in Table 2. Each pixel in the telescope can also be visualized as a "light bucket" which can be overfilled and bled into neighboring pixels, making it necessary to vary exposure lengths to gather the most accurate data possible. Shorter exposures can be used to examine bright stars before their light overflows the buckets in the telescope and longer exposures can be used to examine dimmer stars, which may not have filled the telescope's buckets enough to be observed in shorter exposures.

**Table 1: Objects Imaged by our Team with Dates, Filters, Exposure Lengths, and Number of Photos with same Settings**

Date	Object Imaged	Filter	Exposure Length (s)	Number of Images with these Settings
7/10/23	Bias	N/A	0	13
	Flat	r'	N/A	13
	Flat	g'	N/A	13
	Arcturus	r'	N/A	2
	Arcturus	g'	N/A	2
	M13	g'	180	3
	M13	r'	180	3
	M39	r'	180	1
	M39	r'	300	1
	M39	g'	120	1
	M39	r'	20	1
	M39	g'	300	1

	M39	g'		20		1
	NGC6826	r'		60		1
	NGC6826	g'		200		1
	NGC6826	Blue		60		1
7/11/23	Bias	N/A		0		11
	Flat	N/A	N/A			24
	M13	g'		10		3
	M13	g'		300		1
	M13	r'		10		3
	M13	r'		300		1
	M13	Blue		15		1
	M13	Blue		30		1
	M13	Blue		120		1
	M39	g'		20		1
	M39	g'		120		1
	M39	g'		300		1
	M39	r'		20		1
	M39	r'		120		1
	M39	g'		300		1
	M39	Blue		20		1
	M39	Blue		120		1
7/22/23	Bias	N/A		0		26
	Flat	OIII	N/A			5
	Flat	g'	N/A			5
	Flat	H $\alpha$	N/A			5
	Veil Nebula	OIII		60		3
	Veil Nebula	OIII		180		3
	Veil Nebula	g'		60		3
	Veil Nebula	g'		180		3

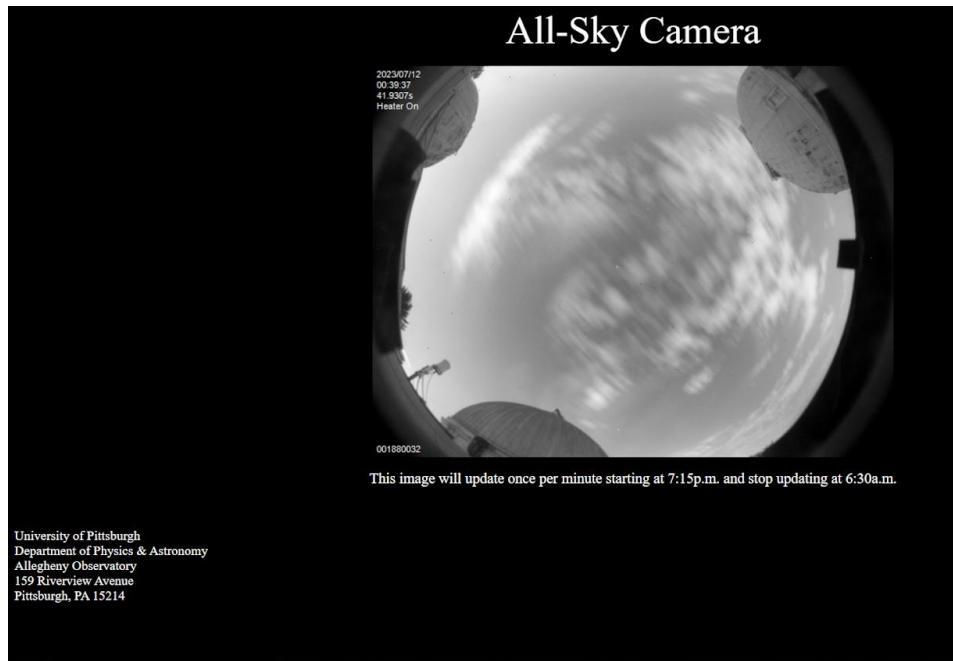
	Veil Nebula	H $\alpha$	60	3
	Veil Nebula	H $\alpha$	180	3

Note: Bias frames can be taken in any filter, but flat frames must be taken in every filter to be used.

**Table 2: Wavelengths That Pass Through Each Filter Used<sup>18</sup>**

Filter:	Wavelength (nm):
r'	550-750
g'	400-600
Blue	300-400
OIII	499-505
H $\alpha$	653-659

It should be noted that the night of 7/10/23 was cut short due to cloudiness, as shown in Figure 8. In addition, on the night of 7/11/23, observation was temporarily interrupted when a loud noise prompted us to take precautionary measures and close the observatory dome for a short period of time.



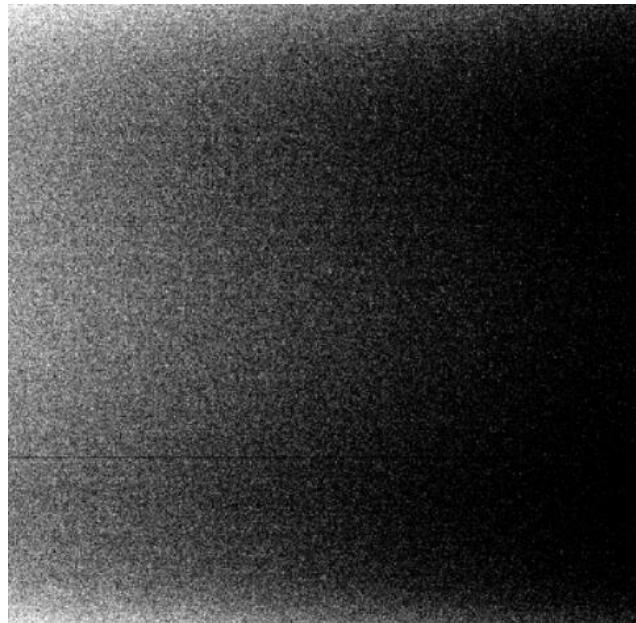
**Figure 8: All-Sky Camera of sky above Allegheny Observatory at 12:39 AM on 7/12/23.**

### III. Data Reduction

To take pictures with the telescope, we must consider any electronic noise that enters the telescope, internal reflections and any unwanted dust and dirt present on the lens. We do this by taking bias and flat frames. These calibration frames are one of the most important steps when measuring the brightness of stars. When determining the age, there are two main components: magnitude and temperature. To find those, the intensities of the stars in the cluster are required. The intensity can be altered depending on extra light or dust in the light images, which is why it is crucial to eliminate those errors before determining the age of the cluster.

#### A. Bias Frames

Bias frames show the electronic noise of the CCD that's unrelated to photons hitting it. This can be very important because without eliminating the bias frames, the final light image can be much brighter than reality. Bias frames are taken without the lens open or pointing at anything, so the picture is in complete darkness. The exposure time for bias frames is near zero, as no outside light is desired. When a bias image is complete, it is observed with counts still in the image. This is exactly what is trying to be removed, since the image was taken in darkness, so there should theoretically be no counts. Typically, several bias images are taken and compiled into one composite image to use for bias subtraction later in the process. In Figure 9 there are counts observed on the left side of the image, which represents electronic noise.

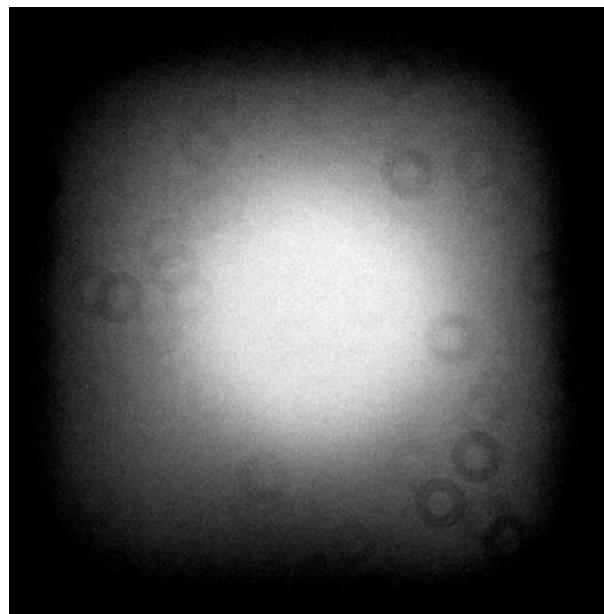


**Figure 9: Bias frame taken with PlaneWave Telescope on 7/10/23.**

#### B. Flat Frames

The next step for calibration frames is taking flat frames. Flat frames are the opposite of bias frames, as they are taken of a white board with lots of light coming in. The purpose of flat frames is to eliminate dust and internal reflections in the light images. In the Allegheny Observatory, there is a large square piece of

plywood that has been painted white. This is the smooth white field of which the telescope takes pictures. There are two lights shining on the board which bring even more light into the image. For flat frames, the images are taken using the same filters that will be used to take the actual light images of celestial bodies. This is because the light is going through the lens, so the different filters can alter how much light comes in. When the flat frames are complete, it is observed with light in the center and a vignette effect with the black. There are also small donut-like objects visible in the image. This is dust observed on the telescope lens, which also must be removed from the final light images. Similar to the bias, several flat images are taken in each filter being used, which are then compiled into multiple flat composite images for each filter. The flat composites are used for flat division later in the process. In Figure 10 there are small rings observed in the image, which represent dust.



**Figure 10: Flat frame taken with PlaneWave Telescope on 7/10/23.**

## B. Composite Images

Now that all the calibration images were taken, they need to be compiled into composite images using the Data Reduction Facility in AstrolImageJ (Appendix A) using a process called “pipelining.” The pipelining process takes all the bias images in one folder and compiles them into one composite image that can be used later. This process repeats for the other calibration frames and for the light image.

The first step of the pipelining process is to create the bias composite. This bias composite will be subtracted from the flat composite and light composite images. Once all bias images are put in one folder, the AstrolImageJ Data Reduction Facility selects all of the images in the folder and compiles them into one composite image. Once this image is created, it can be subtracted from the flat images to make a flat composite image.

For the flat composite image, there is still the extra electronic noise, so it needs to be subtracted. Since the bias composite was already created, all that is needed is to enable bias subtraction in the Data Reduction Facility in AstrolImageJ. This will remove that extra noise from the composite image that will be created.

Similar to the bias composite, all flat images need to be saved to a folder. However, each folder needs to only have one filter in it. There will be several flat composite images made for each filter. Each composite image for the filter will have the bias subtracted from it. Once all of the composites are created, they will later be used for flat division in the light images.

The final step of the pipelining process is to actually create the light images. The end goal with the light images is to be able to get the intensity of the stars to determine the age and to create a three-color image. Since the bias and flat composites were already created, they can now be used for bias subtraction and flat division for the light composite images. Through the same process as the calibration frames, all of the light frames of one filter need to be saved into a folder. There will be multiple folders for each filter. Once all the folders are made, they can individually be processed through AstrolImageJ. Each folder will have the composite bias subtracted and its individual flat composite divided to make it the clearest image possible. Once each folder is processed, there will be individually calibrated images, meaning that the bias is subtracted, and the flat is divided. These images now need to be aligned and stacked together and merged into one single image for each filter. Once this is done there will be light composite images of the celestial object, Figure 11.

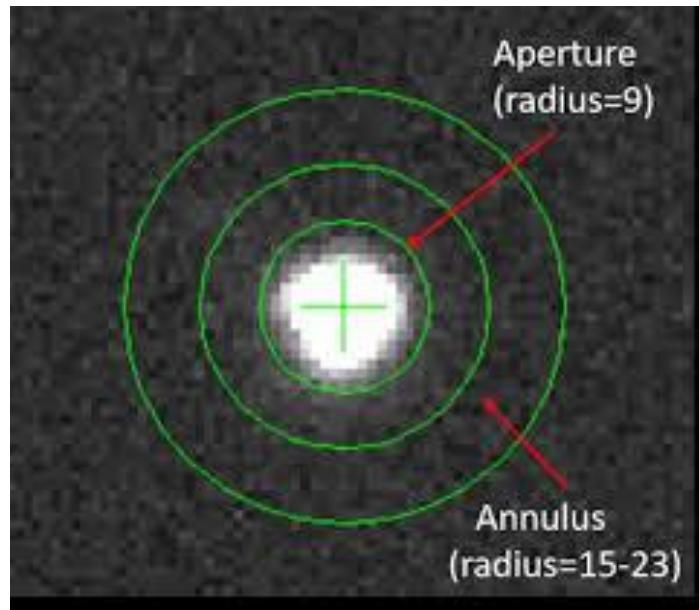


**Figure 11:** Light composite of M13 from PlaneWave Telescope on 7/10/23 and 7/11/23

## D. Photometry

Once the images of an astronomical object are calibrated, there are still a number of calculations that must be done to turn the pictures into usable data. To start this process, images were opened in AstrolImageJ to find the intensities of individual stars. Finding the intensity is done using a multistep process called photometry. First, three circles are placed around the star, dividing it into two main regions: the aperture and the annulus as shown in Figure 12. The purpose of these two regions is to compare the photon count coming from the star to the photon count coming from the background, so the amount of light that a star is

actually emitting can be determined. After the aperture and annulus are placed, AstrolImageJ calculates the total photon count inside the aperture and finds the average photons per pixel of the annulus by counting the photons and dividing them by the area. Next, the photons per pixel of the annulus is multiplied by the area of the aperture, giving the theoretical number of photons in the area of the aperture that would be coming from the background. This amount is subtracted from the total amount of photons in the aperture, leaving only the photons actually coming from the star. AstrolImageJ then performs a series of calculations on this value to yield the intensity of the star.



**Figure 12: Aperture and annulus during the photometry process<sup>19</sup>.**

This process was done roughly 900 times to get the intensities of stars in the globular cluster, with all the calculations being done by AstrolImageJ software. This data was then entered into Google Sheets and Excel to be organized.

## E. Converting to Apparent Magnitude

For this data to be used, it must next be converted from intensities into apparent magnitudes. Intensity is measured with a specific telescope, and is a subjective measurement based on the conditions of the night it is observed. However, the intensity of one star, called a target star, can be compared to the previously known intensity and apparent magnitude of a comparison star, and the apparent magnitude of the target star can be calculated. Because it is calculated in comparison to a star for which the apparent magnitude and intensity are already known, the factors that make the measured intensity of the target star subjective will be accounted for. This is done using the equation

$$m_1 - m_2 = -2.5 \left( \frac{f_1}{f_2} \right) \quad (1)$$

where  $m_1$  is the apparent magnitude of the comparison star,  $m_2$  is the magnitude of the target star, and  $f_1/f_2$  is the ratio of their intensities. By using the values for  $f_1$ ,  $m_1$  (intensity and magnitude of comparison star)

and  $f_2$  (intensity of target star), the magnitude of the target star,  $m_2$ , can be calculated. This equation was used to calculate the apparent magnitudes for all the stars that were analyzed, using data from SIMBAD (Appendix B) to for the comparison star<sup>20</sup>.

## F. Converting to Absolute Magnitude

For this project, the data is used in an H-R diagram, which requires the absolute magnitudes of all the stars in a region. The absolute magnitude of a star is the magnitude that it would have if it were viewed ten parsecs away from the sun. Absolute magnitude is necessary because apparent magnitudes don't account for the distance of a star. As a star gets farther away, the light becomes dimmer; without absolute magnitudes, it would be impossible to differentiate between a star that emits a small amount of light but is close to the observer, and a star that emits a large amount of light but is far away from the observer. The absolute magnitude can be calculated by relating the distance of the star and its apparent magnitude in an equation given by

$$m - M = 5(d) - 5 \quad (2)$$

where  $m$  is the apparent magnitude,  $M$  is the absolute magnitude, and  $d$  is the distance of the star.

Once these calculations were done, there was a lack of particularly dim stars from the data taken with the telescope, so it was supplemented with data from SIMBAD.

## IV. H-R Diagrams for Globular Clusters

There are some unique features in H-R diagrams for globular clusters, an example of which can be seen in Figure 13. Most obvious is the significant deviation from the main sequence as temperature increases. This phenomenon occurs because all the stars in a globular cluster were born at around the same time, but they do not all burn at the same temperature. The hotter a star burns, the quicker it will use up its fuel, meaning a hotter star's lifespan is shorter than its cooler relatives. The hotter stars will enter their giant phase earlier, and that is why this *turn-off point* can be observed in the H-R diagram<sup>21</sup>.

As it turns out, this phenomenon is one of the things that makes globular clusters so interesting to study. By measuring the absolute V magnitude where the turn-off point occurs, it can be found what type of stars have most recently entered their giant phase, and an Isochrone chart (shown in Figure 14) can be used to find which points correspond to which ages<sup>21</sup>.

In general, globular clusters tend to be some of the oldest stellar structures in the universe. For that reason, it is also common for the H-R diagram of a globular cluster to include a *horizontal branch*. These are stars that are so old that they are burning helium instead of hydrogen, causing them to deviate from the main sequence<sup>21</sup>.

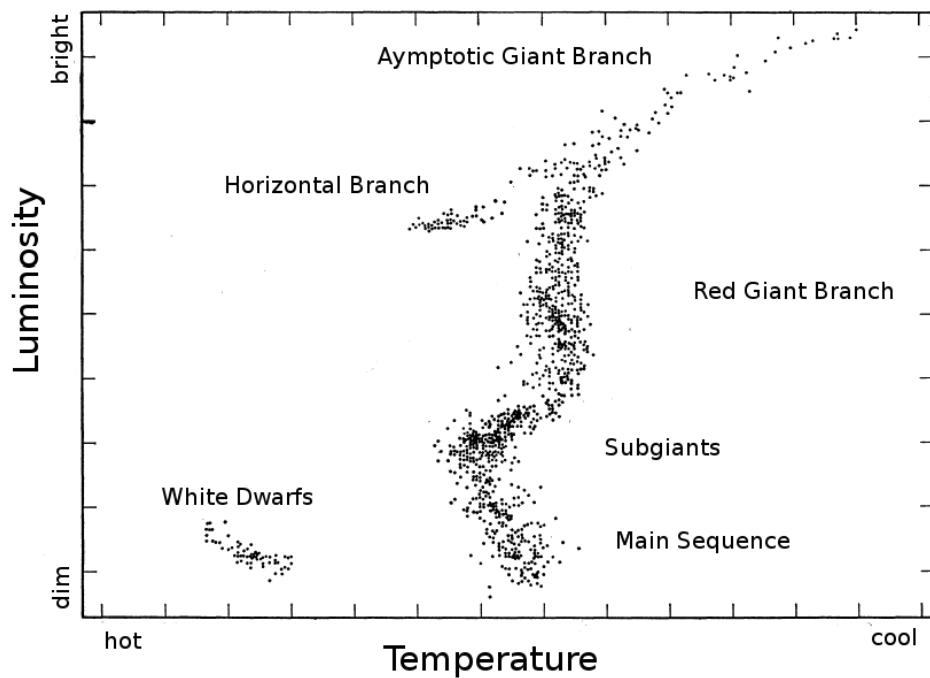


Figure 13: General H-R Diagram for Globular Clusters<sup>22</sup>.

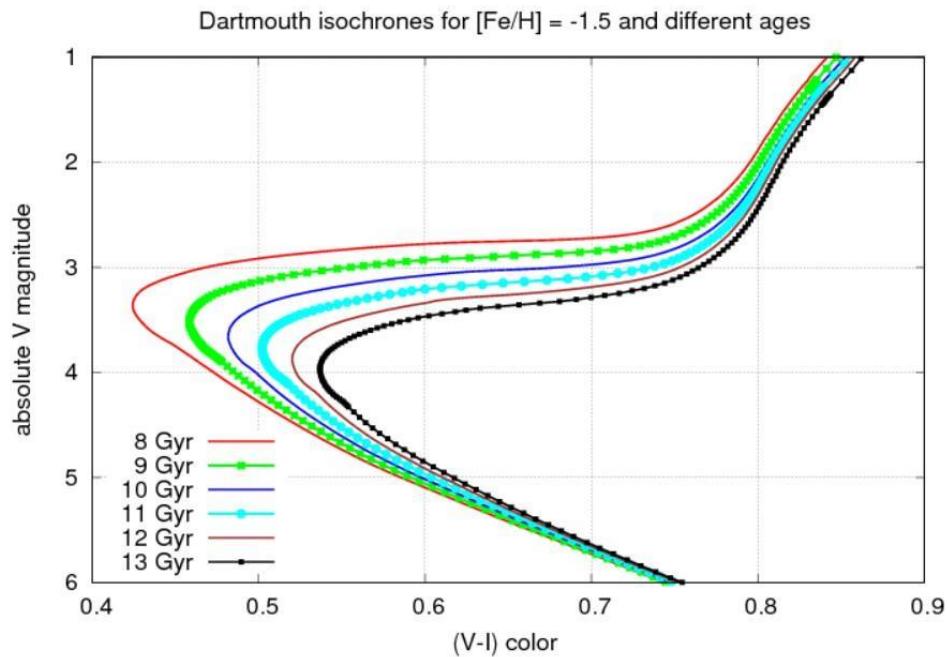
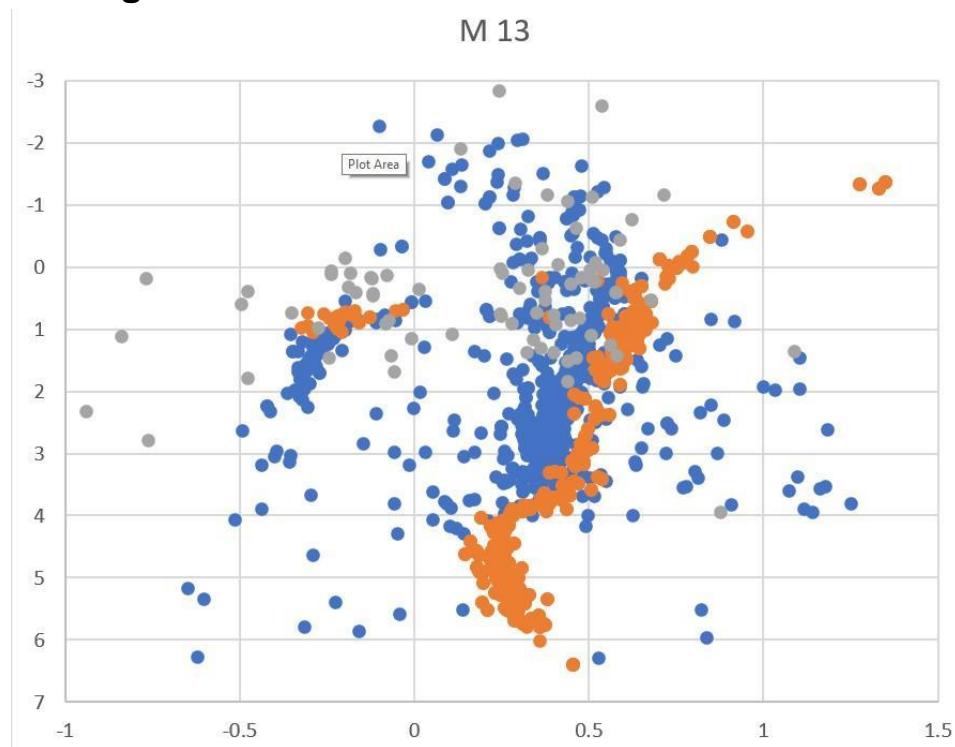


Figure 14: Isochrone Chart

## V. Our H-R Diagram



**Figure 15: H-R diagram of M13.**

Using the data that was collected from the images of M13, an H-R diagram was constructed to determine the age of the globular cluster. The diagram utilized data from 300-second and 10-second-long exposures and used roughly 900 stars altogether. It was supplemented by data from SIMBAD, an astronomical database with basic information about objects outside the solar system. This was done to better define the main sequence of the diagram, especially the fainter stars. Because the M13 globular cluster is so old, its turnoff point is defined mostly by older, fainter stars that were not captured in the images we took. This is another reason why the data used in the H-R diagram was supplemented with SIMBAD.

On the diagram, the y-axis represents the absolute  $g'$  magnitude, while the x-axis represents temperature. One notable aspect of the diagram is the cluster of stars on the upper left side, which is M13's *horizontal branch*. The presence of this branch indicates the old age of M13 and helps support the greater conclusion regarding the age of the cluster that was drawn from the diagram.

### A. Interpreting our H-R Diagram

The H-R diagram above reveals a significant amount of information about the age of the M13 globular cluster. The approximate age can be determined by correlating the absolute  $g'$  value of the turnoff point to an age using an isochrone chart. To do this, the turnoff point was first determined to be at an absolute  $g'$  magnitude of roughly 4.5. This value was then translated into an absolute V magnitude because the isochrone chart used an absolute V magnitude scale. To perform this translation, a series of calculations were performed that culminated in subtracting the absolute  $g'$  magnitude by a value of 1.1. This produced an absolute V magnitude of 3.4. This value was then compared to an isochrone chart, which correlates

absolute V magnitude and metallicity, or the relative ratio of the quantity of other metals and materials to hydrogen, to determine the age of a celestial object.

Figure 14 demonstrates a series of ages for a metallicity of -1.5 depending on an absolute V magnitude. The metallicity of a celestial object does not change after its birth. The metallicity of M13 has been previously determined to be -1.5. The above chart demonstrates several different potential ages and their corresponding curves. The absolute V magnitude of M13 was determined to be 3.4; the peak of the 9 gigayear curve matches this value most closely. Therefore, M13 was determined to be 9 gigayears old. This compares well with the accepted age of 11 gigayears.

## B. Considering Potential Errors

Potential errors and inaccuracies that occurred during this process should be considered. One item that did not impact the integrity of our data but could have been helpful in pinpointing the age of M13, would have been to add error bars on our data. We also did not get an adequate sampling of dim stars, which was impactful because the turnoff point of M13 was mainly defined by older, dimmer stars. Another potential source of error could have been that we did not remove any foreground or background objects, which include several stars and galaxies. The light of these objects had the potential to compromise our data. We also did not correct for any gravitational lensing or redshift effects, but these topics will be discussed later in this paper.

# VI. Other Imagery

When the images are processed, there will be multiple composite images with different filters. These filters can be used to create three color images. Each composite image is slightly different depending on how much light came through the filter. This means that when the images overlap, there will be some layers that are different and some that are the same. When colors are assigned to the different layers, multiple colors come through and a beautiful image is created.

## A. Using FitsLiberator

After creating the composite images, they are saved as FITS (Flexible Image Transport System) File. These composite images can then be uploaded into a program called FitsLiberator (Appendix C). This program allows grayscale scaling to be nonlinear and brings out fine details in order to get the clearest and sharpest image possible. This process gets repeated for all light composites with different filters.

Using PhotoPea, these images can be layered on top of each other. Each layer will get a different color assigned to it based on its wavelength. For example, the shorter wavelengths (400-550 nm) will be blue, the medium wavelengths (550-700 nm) will be green, and the longest wavelengths (700-900 nm) will be red. When all the layers have a different color, a beautiful three-color image will be created. A few extra steps to make the image clearer would be to add curves into the layer. These curves will change the amount of color shown on the screen, so there can be more or less of one color to make the image clearer or brighter.

## B. Messier 13 Three-Color Image

For M13, our color image was very purple. This is because our main images were from blue and red wavelengths. Our third filter was more of a supplemental filter, so the major colors are still blue and red. We applied the same imaging process as described above to create our three-color image of M13 as shown in Figure 15.

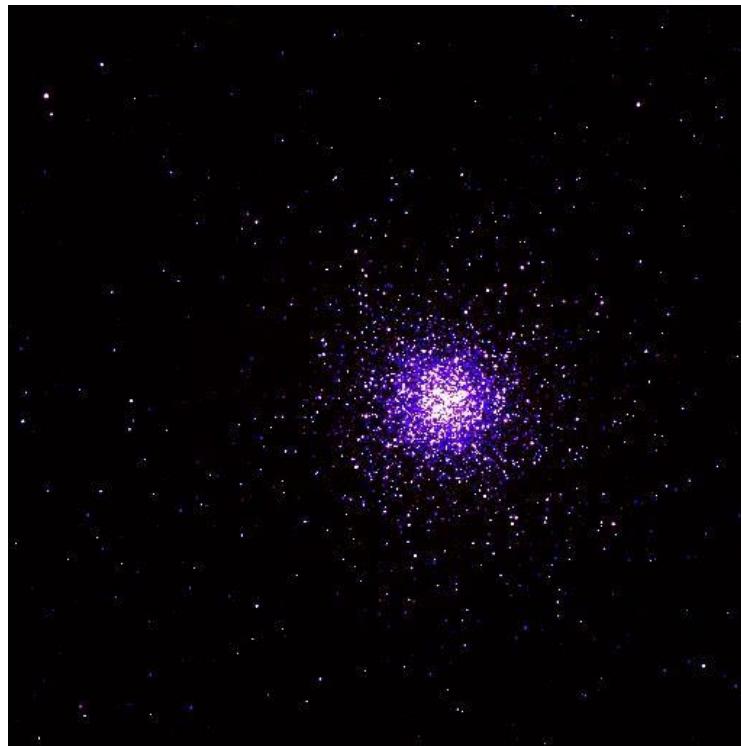
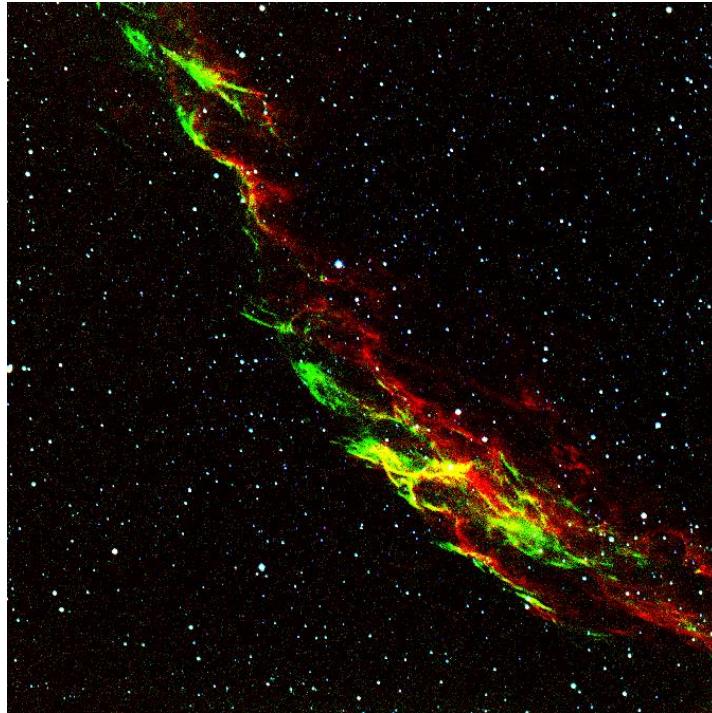


Figure 15: Three-color image of Messier 13 globular cluster.

## C. Veil Nebula Three-Color-Image

On a separate night, we took images of the Veil Nebula. We did not use these images to determine the physical properties of the nebula but to generate our own three-color image of a nebula in space. The three filters that we used for this process were g', OIII, and H-Alpha. The Veil Nebula is an emission nebula and is very bright in the emission lines of Hydrogen and doubly ionized Oxygen. We still went through the pipelining process with bias and flat frames, and then created our light images in the three different filters. When we generated our three-color image, the nebula was mainly green and red. This was because the OIII and H-Alpha filters let in a smaller amount of light, so the nebula showed up very bright. The g' filter lets in more light, but that also means that the images are less clear. As a result, the g' filter left mainly stars in the image and didn't show much of the nebula. When layering the images all together, we were able to get a very clear image of the Veil Nebula, with predominantly red and green colors as shown in Figure 16.



**Figure 16: Three-Color-Image of the Veil Nebula taken with the PlaneWave Telescope on 7/22/23**

#### D. Hubble Legacy Archive

Another project that we worked on was taking images from the Hubble Legacy Archive (Appendix D) and making three color images out of them. The Hubble archive has gray-scale images of several nebulas, galaxies, globular clusters, and other galactic structures. These gray-scale images have different filters for each image that we can download and make our own images out of them. We used the same software, FitsLiberator, to adjust the gray-scale images from the archive. Then we put all of the images with the different filters into PhotoPea to make our three-color images. We did this for several nebulas including the Eagle Nebula, Orion's Nebula, Lagoon Nebula, Omega Nebula, and ESO Planetary Nebula as shown in Figure 17.



**Figure 17: Three Color Images Created Using Grayscale Images from Hubble Legacy Archive**  
**Going clockwise from top left: Eagle Nebula, Lagoon Nebula, Omega Nebula, ESO Planetary Nebula, Orion's Nebula**

## VII. Redshift

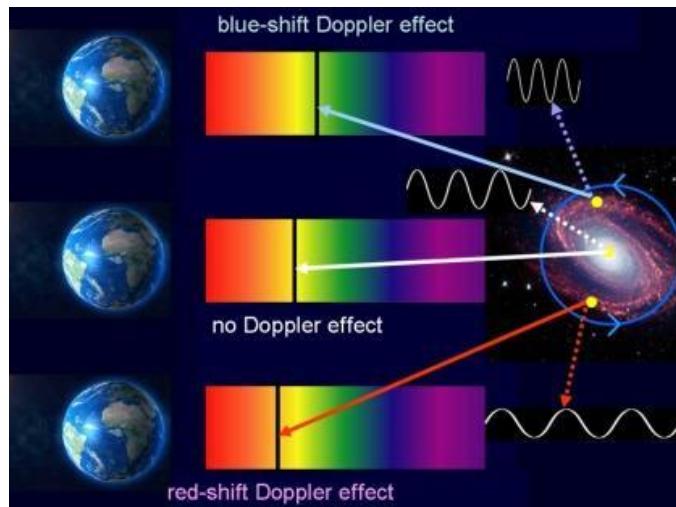
### A. Defining Redshift

'Red shift' is an important calculation used by astronomers that helps compare the distance of faraway galaxies, as well as offering insight to the chemical composition of distant celestial objects<sup>23</sup>. There are three kinds of 'redshift': astronomical, gravitational, and doppler. This experiment only requires the use of astronomical redshift and doppler redshift, as gravitational red shift does not pertain to the scope of the experiment's research goals. The idea of redshift is better understood using the context of what some might have heard of: the Doppler effect. Doppler effect transpires when a source of sound moves relative to an observer. It was discovered and named after Christian Andreas Doppler, an Austrian mathematician who discovered the frequency of sound wave changes<sup>24</sup>. Numerous examples of the Doppler effect include the changing of pitch of police and ambulance sirens, or train whistles and racing car engines as they pass by.

### B. Doppler Redshift

Considering the fact that light behaves similarly to a wave, the light from a luminous celestial object (in this case globular clusters) undergoes a Doppler-like shift if the source is moving relative to the observer's perspective. Edwin Hubble, the American astronomer who discovered the universe is expanding in 1929, has motivated astronomers to continue their study of redshifting by assuming that *most* other galaxies are

moving away from Earth. The wavelength of light emitted by these galaxies becomes stretched the farther it moves away from the observer, causing it to appear as 'redshifted'.



**Figure 18:** As a celestial object emitting light moves closer to the observer, the frequency of the wavelength grows compressed, and will appear blue. However, if the object were to be moving away from the observer, the wavelength would stretch, and it would appear red to the human eye<sup>25</sup>.

The red shift of a specific globular cluster can be easily measured by comparing its respective spectrum with a reference laboratory spectrum. At well-known wavelengths, absorption lines can be used as a reliable indicator to determine the redshift of the receding astronomical objects by measuring the location of these lines in astronomical spectra.

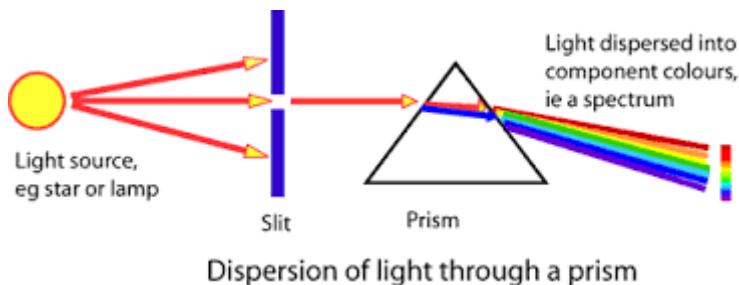
When the velocity of the astronomical object is comparably small to that of the speed at which its light waves propagate, the magnitude of either the redshift or the blueshift is referred to as variable 'z' and is given by

$$z = \frac{\Delta\lambda}{\lambda_{rest}} = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}} = \frac{v}{c} \quad (3)$$

where  $\Delta\lambda$  is the change in wavelength of the light,  $\lambda_{rest}$  is the wavelength of the light if the object was stationary,  $\lambda_{obs}$  is the observed wavelength of the light,  $v$  is the velocity of the moving object, and  $c$  is the speed of light in a vacuum.

### C. Absorption Lines & Spectra

Astronomers are then able to observe how redshift and blueshift transpire over time by using a high-resolution prism-like instrument also known as a spectrograph. The spectrograph separates the incoming wavelengths of light into an array of colors, also known as the spectra.



**Figure 19: The process of using a spectrograph**  
Source: Australia Telescope National Facility<sup>26</sup>.

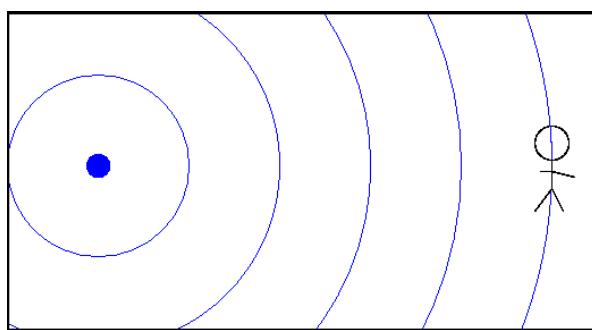
Within every star, there are atoms that absorb light at specific wavelengths, and these absorptions appear as dark lines in the different colors of the star's spectrum, better known as spectra. Researchers use the shifts in these lines as convenient markers to measure the size of a Doppler shift.

Knowing the specific frequencies of Hydrogen lines, astronomers study light emitted by these stars. You can substitute Hydrogen with different elements or molecules to get completely different distinct patterns of absorption lines, also known as the distinct absorption fingerprints. From these distinct absorption fingerprints, astronomers are able to discover the chemical makeup of stars and galaxies<sup>26</sup>.

### C. Relativistic Doppler Effect

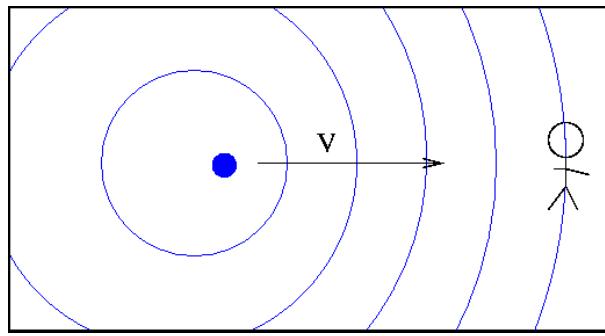
A more mathematical approach to the Doppler effect leads one to understand the special relativity behind 'The Relativistic Doppler Effect' equation. This formula begins with the general understanding that waves of any sort- whether sound or light- that are emitted at some frequency by a moving object are perceived differently by a stationary observer.

When both the source of the wave as well as the observer are both stationary, the observer would 'see' waves of frequency, ( $v$ ) or wavelength, ( $\lambda$ ).



**Figure 20: A stationary source of wavelength (left) and a stationary observer (right) as the waves emitted by the stationary source reaches the observer**<sup>27</sup>.

However, if the source was moving towards the observer, then the frequency which is perceived by the observer ( $v'$ ) **would** be faster than the actual emitted frequency, thus the understood wavelength from the perspective of the observer ( $\lambda'$ ) **is** shorter than the emitted wavelength.



**Figure 21:** This source of wavelength (left) as it moves towards the observer (still stationary) at a certain velocity ( $v'$ ) would cause the frequency to be perceived as higher, while the wavelength would be shorter<sup>27</sup>.

Astrophysicists can connect the emitted and observed frequencies and wavelengths of Doppler shifted light by equations given by

$$f' = \frac{f}{1 \mp \frac{v}{c}} \quad (4)$$

and

$$\lambda' = \lambda \left(1 \mp \frac{v}{c}\right) \quad (5)$$

where  $f'$  and  $\lambda'$  are the frequency and wavelength respectively perceived by the observer,  $f$  and  $\lambda$  are the frequency and wavelength respectively emitted from the object,  $v$  is the velocity of the object, and  $c$  is the speed of light in a vacuum.

It should be noted that the shift in wavelength for a moving source moving *towards* the observer is directly inverse to the shift for a source moving *away* from the observer.

By using the postulates of Einstein's Theory of Special Relativity, astrophysicists can derive a different relationship between frequency and motion<sup>5</sup>. A simple way to understand it is to note that the time on a moving source runs slow by factor of ' $\gamma$ ' so that it appears to a stationary observer that it would take a bit longer to create each wave crest. Therefore, the rate at which new wave crests appear (frequency) is slowed down by  $\gamma$ . The equation is given by

$$v' = \frac{v}{1 + \frac{v}{c}} \frac{1}{\gamma} \quad (6)$$

which, after doing some algebra, can be simplified to

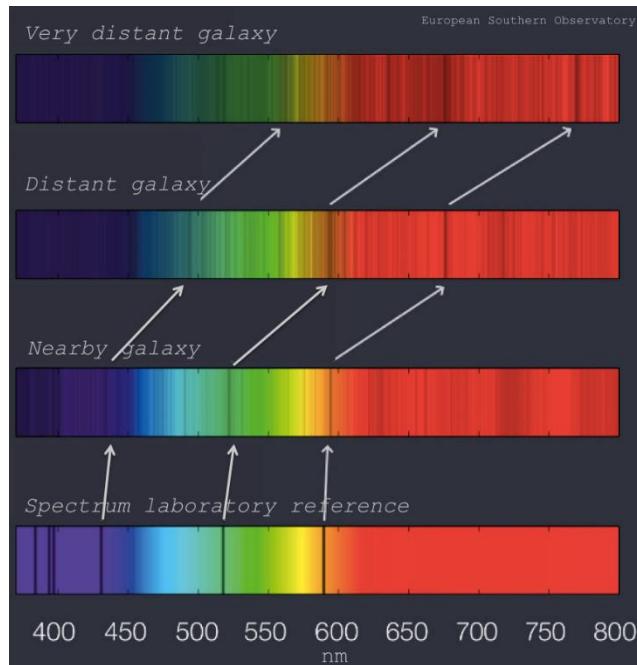
$$v' = v \frac{\sqrt{1-\frac{v}{c}}}{\sqrt{1+\frac{v}{c}}} \quad (7)$$

where velocity is negative, the source and object are approaching each other. Where the velocity is positive, the source and object are retreating from each other. Using this equation, astronomers can now measure the speed at which stars are either approaching or receding from the observer.

## E. Astronomical Redshift

The red shifts observed in distant objects are not entirely due to Doppler redshifting, but are also affected due to the expansion of the universe. While Doppler redshifts stem from the relative motion of the source of frequency and the observer through space, astronomical redshifts take into account the expansion of space itself<sup>6</sup>. This means that although two astronomical bodies can be stationary in space, they will still be subject to astronomical red shifting due to the rate at which the universe is growing.

In the early 20th century, the astronomer Edwin Hubble observed that the spectra of distant galaxies were significantly redshifted. Edwin Hubble determined that this drastic shift in the spectrum is reliant on velocity in which these galaxies were receding, which was also proportional to how far they are relative to the observer.



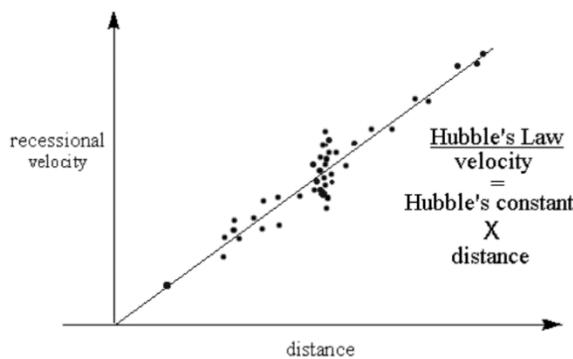
**Figure 21:** This shows the difference in spectra among nearby galaxies, distant galaxies, and very distant galaxies. As previously stated by Hubble's discovery, the farther the galaxy is relative to the observer, the farther it is redshifted<sup>28</sup>.

Hubble's conclusion that the velocity in which the galaxies are receding from the observer is equal to the distance is mathematically shown in Hubble's Law.

## F. Hubble's Law

Hubble's law, as an example, can be represented graphically by creating a plot of multiple different observed galaxies. The x-axis of the chart depicts the distance, "d", of each galaxy- typically in megaparsecs. On the y-axis, astronomers plot the velocity, "v", at which the celestial body is receding relative to the observer. This is usually measured in kilometers per second, km/s.

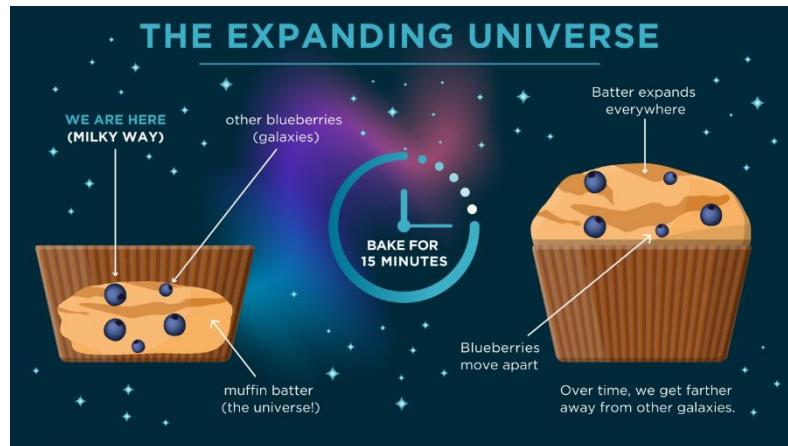
The slope of the line created is called, ' $H_0$ ', pronounced H-naught, also known as the Hubble constant which has a value of roughly 68.8 km/sMpc.



**Figure 22:** This shows Hubble's law depicted graphically as it created a plot of each observed galaxy using the velocity in which it is receding from the observer (recessional velocity) over the distance that it is relative to the observer (measured in megaparsecs)<sup>29</sup>.

From this law, he, as well as a few other astronomers, were able to conclude several important facts pertaining to the nature of this Universe. These conclusions stem from the claim that all galaxies appear to be receding from a specific point in space. This stems from the fact that the entire universe is constantly expanding; therefore 3-D space-time itself is stretching apart. Regardless of the vantage point, all galaxies appear to be moving away<sup>30</sup>. Hubble deduced that the farther away the galaxy is, the higher velocity in which it is receding. Additionally, the rate of expansion of the Universe is steadily growing faster than how it used to when the Universe was young.

This expansion of space-time is more simply depicted using the analogy of baking a blueberry muffin. The batter will represent the Universe and the blueberries are all which reside within it. Although the blueberries themselves remain stationary in their own perspective, as the batter continues to bake, it causes the blueberries to move as well. The blueberries furthest from each other move away the quickest.



**Figure 23:** A drawn example of the blueberry muffin analogy used to explain the expansion of the universe<sup>29</sup>.

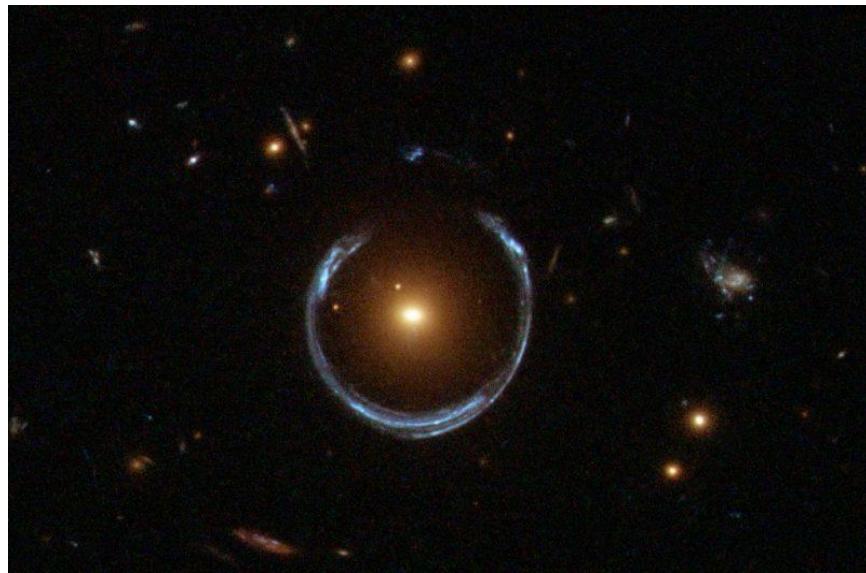
The first conclusion made by Hubble and a few other astronomers was that if the Universe is always moving apart, then if one were to rewind to the early Universe, everything would be compact into one small point. This single hot point then rapidly expanded outward: the moment of the Big Bang.

If astronomers were to look at galaxies of ranging distances and observe how fast they are receding due to the expansion of space-time, they would determine that there was a varying rate for the expansion of the universe<sup>30</sup>. Regardless of the varying speed, astronomers can then work backwards to further understand when the movement began. From Hubble's observations, astronomers can confidently conclude how long ago the Big Bang occurred, and therefore estimate the age of the universe.

## VIII. Gravitational Lensing

Gravitational lensing is a phenomenon that is often observed when looking into deep space. It occurs when a massive celestial body causes a curvature of spacetime large enough to bend light. This can occur with any massive enough object, including globular clusters such as M13. The body causing the light to bend is referred to as the gravitational lens. It was first proven during the solar eclipse of 1919, when astronomers witnessed the sun bending the light of background stars by the amount predicted by Albert Einstein.

The phenomena works in conjunction with Einstein's theory of relativity. This theory states that space and time are fused together in a body called spacetime that makes up the "fabric" of the universe. One of the major ideas of this theory is that massive objects distort spacetime in a manner that can impact the path of anything traveling in the impacted region, including light itself.



**Figure 24:** This image, taken by the Hubble Space Telescope's Wide Field Camera 3, depicts a luminous red galaxy with a strong gravitational impact that warps the light of a more-distant blue galaxy<sup>31</sup>.

Gravitational lensing is useful because it helps astronomers detect and understand the presence of dark matter in the foreground of galaxies. Although dark matter makes up as much as 85% of the total mass in the universe, it emits no light, and therefore is challenging to detect. However, astronomers can study gravitational lensing to discern the location of dark matter based on its impact on the light around it. In this way, gravitational lensing contributes to humanity's understanding of the universe<sup>31</sup>.

## IX. Conclusion

Globular clusters are a very important and interesting part of the universe. They are compact, spherical groups that consist of millions of stars, and move through the universe as a whole. Because the stars within globular clusters are so old, there are a number of insights that can be gained only from these clusters. While their origins are not fully understood, they provide a way to learn more about the distribution of dark matter, the age and history of the universe, and different features of galaxies.

In this project, the age of the globular cluster M13 was determined through a long process involving gathering data with a telescope, processing the data, and plotting an H-R diagram that was then compared to an Isochrone chart. Images with varying exposure lengths and different filters were taken with a 24" PlaneWave telescope, which were then calibrated using bias and flat frames that were also taken on the telescope. AstroImageJ software was used to find the intensities of individual stars in the cluster through photometry. This data was then converted from intensities to apparent magnitudes, then to absolute magnitudes. The absolute magnitudes were then plotted on the H-R diagram. The H-R diagram plots the absolute magnitude on the y-axis and the value of the blue light minus visual (green) light, which represents the temperature. This was then compared to an Isochrone chart, which shows how the turnoff point on the H-R diagram indicates the age of the cluster.

Through this comparison, M13 is estimated to be 9 gigayears old. The effect of gravitational lensing and redshift is discussed, as it pertains to potential errors in the data and results. After the data is used for determining the age of the cluster, the pictures taken with the telescope were then colored and layered to create 3-color images of various celestial objects. Along with these pictures, other 3-color images were created using grayscale images from the Hubble Legacy Archive. In conclusion, globular clusters are some of the most interesting celestial objects to study as they give insights into the nature of the universe as a whole.

## X. Appendix

### A. AstrolImageJ

AstrolImageJ is ImageJ with astronomy plugins and macros installed. It includes tools based on the Göttingen ImageJ astronomical resources with additions we find useful.

### B. Simbad

The SIMBAD astronomical database provides basic data, cross-identifications, bibliography, and measurements for astronomical objects outside the solar system. SIMBAD can be queried by object name, coordinates, and various criteria.

### C. FitsLiberator

The ESA/ESO/NASA FITS Liberator is a free software program for processing and editing astronomical science data in the FITS format to reproduce images of the universe.

### D. Hubble Legacy Archive

The Hubble Legacy Archive (HLA) is designed to optimize science from the Hubble Space Telescope by providing online, enhanced Hubble products and advanced browsing capabilities. The HLA is a joint project of the Space Telescope Science Institute (STScI), the Space Telescope European Coordinating Facility (ST-ECF), and the Canadian Astronomy Data Centre (CADC)

## XI. References

<sup>1</sup>Information@eso.org. "Open Cluster." *ESA/Hubble | ESA/Hubble*, esahubble.org/wordbank/open-cluster/. Accessed 3 Aug. 2023.

<sup>2</sup>Gianopoulos, Andrea. "Discoveries - Hubble's Star Clusters." NASA, 17 Nov. 2022, [www.nasa.gov/content/discoveries-hubbles-star-clusters](http://www.nasa.gov/content/discoveries-hubbles-star-clusters).

<sup>3</sup>"Halo: Cosmos." *Halo | COSMOS*, astronomy.swin.edu.au/cosmos/h/halo. Accessed 3 Aug. 2023.

<sup>4</sup>*Globular Cluster Systems - California Institute of Technology*, ned.ipac.caltech.edu/level5/Ashman/Ashman1.html. Accessed 4 Aug. 2023.

<sup>5</sup>Mohon, Lee. "Galactic Center." NASA, 22 July 2019, [www.nasa.gov/mission\\_pages/chandra/images/galactic-center.html](http://www.nasa.gov/mission_pages/chandra/images/galactic-center.html).

<sup>6</sup>*Globular Cluster Formation and Evolution in the Context of Cosmological ...*,  
royalsocietypublishing.org/doi/10.1098/rspa.2017.0616. Accessed 3 Aug. 2023.

<sup>7</sup>“Star Clusters.” *Star Clusters* | Center for Astrophysics, pweb.cfa.harvard.edu/research/topic/star-clusters. Accessed 3 Aug. 2023.

<sup>8</sup>Trevor. “What Is a Globular Cluster?: Pictures, Facts, & Best Ones to Observe.” *AstroBackyard*, 26 Oct. 2022, astrobackyard.com/globular-clusters/.

<sup>9</sup>[Information@eso.org](mailto:Information@eso.org). “The Hubble Tuning Fork - Classification of Galaxies.”  
www.Spacetelescope.org, esahubble.org/images/heic9902o/. Accessed 4 Aug. 2023.

<sup>10</sup>*Astronomy 505*, astronomy.nmsu.edu/nicole/teaching/ASTR505/lectures/lecture27/slide01.html.  
Accessed 3 Aug. 2023.

<sup>11</sup>*Spiral Structure in Galaxies: The Classification of Galaxies* - IOPscience,  
iopscience.iop.org/chapter/978-1-6817-4609-8/bk978-1-6817-4609-8ch2.pdf. Accessed 4 Aug. 2023.

<sup>12</sup>“Types of Galaxies.” NASA, 10 Apr. 2023, universe.nasa.gov/galaxies/types/.

<sup>13</sup>*Grard de Vaucouleurs' Atlas of Galaxies*, cseligman.com/text/devaucouleurs.htm. Accessed 3 Aug. 2023.

<sup>14</sup>*The Hertzsprung-Russell (HR) Diagram*, spiff.rit.edu/classes/phys230/lectures/hr/hr.html. Accessed 3 Aug. 2023.

<sup>15</sup>Sloan Digital Sky Survey (SDSS) Filters. - Researchgate, [www.researchgate.net/figure/Sloan-Digital-Sky-Survey-SDSS-filters\\_fig5\\_235663628](http://www.researchgate.net/figure/Sloan-Digital-Sky-Survey-SDSS-filters_fig5_235663628). Accessed 4 Aug. 2023.

<sup>16</sup>“Hertzsprung Russell Diagram.” *AstroPages* | HR Diagram | Western Washington University,

<sup>17</sup>*History of Allegheny Observatory*, sites.pitt.edu/~aobsvtry/history.html. Accessed 3 Aug. 2023.

<sup>18</sup>“Mass, Luminosity, and the HR Diagram.” Mass and Luminosity,  
[pages.uoregon.edu/soper/Stars/hrmass.html](http://pages.uoregon.edu/soper/Stars/hrmass.html). Accessed 4 Aug. 2023.

<sup>19</sup>Conti, Dennis M. *Exoplanet Observing*, astrodennis.com/Guide.pdf. Accessed 4 Aug. 2023.

<sup>20</sup>“17.1 the Brightness of Stars - Astronomy.” OpenStax, openstax.org/books/astronomy/pages/17-1-the-brightness-of-stars. Accessed 3 Aug. 2023.

<sup>21</sup>*Durham E-Theses Globular Cluster Systems and Their Implications Of ...*,  
etheses.dur.ac.uk/2427/1/2427\_438.pdf. Accessed 4 Aug. 2023.

<sup>22</sup>“The Life Cycle of Stars.” *The Life Cycle of Stars* | Max Planck Institut Für Sonnensystemforschung,  
www.mps.mpg.de/de/sage/life-cycle-stars. Accessed 3 Aug. 2023.

<sup>23</sup>“What Do Redshifts Tell Astronomers.” *EarthSky*, 2 Jan. 2022, earthsky.org/astronomy-essentials/what-is-a-redshift/.

<sup>24</sup>“What Is ‘Red Shift?’” *ESA*, www.esa.int/Science\_Exploration/Space\_Science/What\_is\_red\_shift.  
Accessed 3 Aug. 2023.

<sup>25</sup>“How Is the Doppler Effect Related to Understanding the Universe?: Socratic.” *Socratic.Org*, 26 Aug. 2015, [socratic.org/questions/how-is-the-doppler-effect-related-to-understanding-the-universe](https://www.socratic.org/questions/how-is-the-doppler-effect-related-to-understanding-the-universe).

<sup>26</sup>“Obtaining Astronomical Spectra- Spectrographs.” *Obtaining Astronomical Spectra- Spectrographs*, 7 June 2023, [www.atnf.csiro.au/outreach/education/senior/astrophysics/spectrographs.html](https://www.atnf.csiro.au/outreach/education/senior/astrophysics/spectrographs.html).

<sup>27</sup>“The Relativistic Doppler Effect.” *Apache2 Ubuntu Default Page: It Works*, [spiff.rit.edu/classes/phys314/lectures/doppler/doppler.html](https://spiff.rit.edu/classes/phys314/lectures/doppler/doppler.html). Accessed 3 Aug. 2023.

<sup>28</sup>“Astronomical Redshift.” *Imaging the Universe - The University of Iowa*, [itu.physics.uiowa.edu/labs/advanced/astronomical-redshift](https://itc.physics.uiowa.edu/labs/advanced/astronomical-redshift). Accessed 3 Aug. 2023.

<sup>29</sup>Staff, Prisha Singh- CuriouSTEM. “Theories in Physics: Hubble’s Law.” *CuriouSTEM*, 11 Nov. 2020, [www.curiousstem.org/stem-articles/theories-in-physics-hubbles-law1](https://www.curiousstem.org/stem-articles/theories-in-physics-hubbles-law1).

<sup>30</sup>Warren, Sasha. “The Hubble Constant, Explained.” *University of Chicago News*, [news.uchicago.edu/explainer/hubble-constant-explained](https://news.uchicago.edu/explainer/hubble-constant-explained). Accessed 3 Aug. 2023.

<sup>31</sup>“What Is Gravitational Lensing?” *EarthSky*, 8 Sept. 2021, [earthsky.org/space/what-is-gravitational-lensing-einstein-ring/](https://earthsky.org/space/what-is-gravitational-lensing-einstein-ring/).



# Light: Wave, Particle or Something Else?

Kofi Anokye and Selina Lin

## Abstract

Light is a peculiar phenomenon of nature, subjecting it to a wide variety of theories over the past millennia. The purpose of this research was to gain a better understanding of light and its properties. In our first experiment, we examined the wave nature of light by performing a double-slit experiment with two different slit separations and then measuring the resulting interference pattern. Using previously determined equations based on the wave theory, we were able to accurately predict the distance between the slits; therefore, our results were consistent with the wave theory. In a related experiment, we quantified the brightness of the interference pattern to verify that the visual pattern aligns with our measurements. Finally, we investigated the particle nature of light by observing the photoelectric effect. We established that the frequency of the incoming light is positively related to the kinetic energy of the emitted electrons, a finding that is consistent with Einstein's particle theory. After demonstrating both the wave and particle natures of light, we contend that light can act as a particle or a wave depending on how it is observed. However, we ponder whether there is an underlying characteristic that can explain both natures as a whole.

## I. Introduction

### A. Background

What is a particle or a wave? A particle means many different things to various physicists—the variegated definitions all capture some element of the nature of particles. Here, we posit that a particle is a pointlike object, a mathematical model with all its energy concentrated at a point.

Waves, on the other hand, are continuous (propagating) disturbances to an equilibrium. The oscillating motion of a spring-mass system in classical physics (a simple harmonic oscillator) does an excellent job of illustrating this. When one pulls on a spring (applies a force) attached to some mass, the position of this system over time results in a wave pattern as the spring goes back and forth, being drawn to equilibrium, pushing past it, being drawn again and pushing past it and so on. If taken as a wave, light would be the oscillation of electric and magnetic fields perpendicular to each other and the direction of travel.

#### 1. History

Light, being a fundamental yet illusory part of human life, has lent it many interpretations over millennia. We recognize that it has something to do with vision because a lack of it precludes sight, but how this was the case was thought of in quite distinctive ways.

Most ancient Greeks thought that a visual fire emanated from our eyes, enabling us to see—a "visual hand", as put by Hipparchus, a Greek astronomer, around 300 BC. In ancient India, certain schools of Hindu philosophy, like the Nyaya and Vaisheshika, considered light or fire among the five elements. There was fire from the eye and fire from the heavenly bodies—both pondered as "streams of atoms". Ibn al-Haytham or Alhazen, a Muslim astronomer and mathematician, in *Kitab al-Manazir* or *The Book of Optics*, contended that light must play a critical role in vision because when light intensity increased, the clarity of vision

improved; the intensity at the extreme was visually disorienting; and after-images occurred after some interaction with light. It was not ocular rays, but "[t]he direct or indirect illumination of the eye" that resulted in vision<sup>1</sup>.

In a 1672 paper in the *Philosophical Transactions of the Royal Society* (a publication of Britain's foremost scientific society), English natural philosopher Isaac Newton reported the results of a prism experiment. He had allowed a circular beam of sunlight to pass through a triangular glass prism, resulting in a "vivid and intense" production of colours in a rectangular form<sup>2</sup>. He had experimental results which disproved the then-popular laws of refraction (of the passage of light between mediums), which told him to expect the colours in a circular form since he utilized a circular beam<sup>2</sup>. In describing the light rays in that experiment, he compares them to tennis balls, suggesting that light was composed of particles<sup>3</sup>. For Newton, light's linear motion indicated that light was the radiation of particles or "corpuscles" from a source<sup>4</sup>.

In the early eighteenth century, Dutch natural philosopher Christian Huygens' wave theory of light contrasted with Newton's corpuscular theory. To Huygens, light, like sound, travelled through a medium, and the vibrations or "impressed motion" of this aether medium caused light<sup>1</sup>. Newton's notoriety made the particle theory the preponderant conception of light among scientists, but substantial evidence still existed for the wave nature of light.

The most significant bolsterer of the wave theory of light was British natural philosopher Thomas Young's double slit experiment. Following Newton, in the macroscopic world, light could not bend around an object like a wave would because it was rectilinear—always travelling in straight lines. Young's experiment, in the early 1800s, contradicted this notion, however. When light passed through a narrow slit, it bent and spread out like a wave would. Newton's particle theory could not explain this phenomenon, so by the mid-1800s, when the wave theory had collected more undergirding evidence, it had become widely accepted.

Around the 1880s, another contradiction emerged—an observation had conflicted with the predominant wave theory of light. This discovery was the photoelectric effect—the emission of electrons from metal when light strikes the metal's surface. German physicist Philipp Lenard had separated white light into component colours using a prism and separately focused each colour onto a metal surface to observe the effect further. To his surprise, electrons emerged with different amounts of energy depending on the colour or wavelength of light (with shorter wavelengths resulting in more energetic electrons). Moreover, increasing the intensity of the light did not affect the energy of the emitted electrons, though the number of electrons that escaped from the metal surface increased. Therefore, contrary to a wave nature, the wavelength, not the amplitude of the light, affected the energy of the liberated electrons.<sup>5</sup>

In 1905, German-born physicist Albert Einstein, to explain Lenard's observed phenomenon, formulated a new particle theory of light. He proposed that each particle of light, or photon, contains a specific amount of energy that depends on the frequency of the light. When these photons collide with electrons, they transfer their energy into the electrons.<sup>6</sup>

The wave and particle properties of light have been confirmed countless times; however, the theories taken separately cannot explain all observed light phenomena. Hence, there still exists some ambiguity as to the true identity of light. Which theory is true? Can both approaches be valid? Or is there another explanation?

## II. Experiments

### A. Thomas Young's Double Slit Experiment

#### 1. Theory

To observe the wave properties of light, we replicated Young's double-slit experiment, using a double slit and a 40-slit plate which we will refer to as Slit A and Slit B respectively. Relying on uninhibited intuition, one may assume that bright spots equal to the number of slits would be illuminated on the sheet. In this scenario, Slit A would give us two bright lines, as shown in Figure 1a. There is instead an array of bright spots. The brightest spot is in the middle, and proceeding towards either side, a symmetry of dimmer and dimmer bright spots separated by dark portions, as shown in Figure 1b.



**Figure 1: a. Intuition (If Light Were a Particle). b. Experimentation**

Per the wave theory of light, we can explain the phenomenon with constructive and destructive interference. Constructive interference is when the amplitudes of waves on coinciding paths add up at each point. At the extreme (pure constructive interference), provided the wavelengths and amplitudes of the light waves are the same, when the peaks of each wave line up, the amplitude of the resulting wave doubles (Figure 2a). When one wave is shifted by half a wavelength, the peak of one wave aligns with the trough of the other, resulting in complete destructive interference—the waves cancel each other out (Figure 2b). When the peaks of the waves line up, they are considered “in phase”; otherwise, the waves are considered “out of phase”.<sup>7</sup>

Conceding to the wave theory, when the light waves travelled through the narrow slits, they diffracted—bended upon encountering an aperture bordered by obstacles—and spread out, causing the light waves to overlap and interfere with each other, resulting in the characteristic pattern. At bright spots, the light experiences constructive interference and at dark spots, destructive interference.

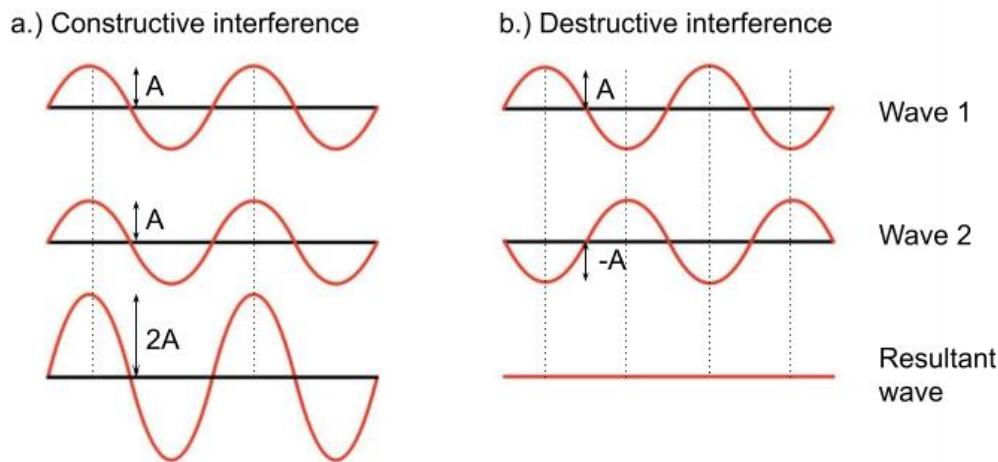


Figure 2: Illustration of Interference

## 2. Quantitative Analysis

To explain this phenomenon mathematically, we can consider two waves travelling to the same point on the screen, as shown in Figure 3. The two waves must travel different distances to reach the given point; therefore, one wave is shifted a certain distance,  $x$ , compared to the other. If this shift is an integer multiple of one wavelength, the waves will still be in phase, and a bright spot will appear at the given point. On the other hand, if one wave is shifted half a wavelength relative to the other, the waves will experience destructive interference and a dark spot will appear.

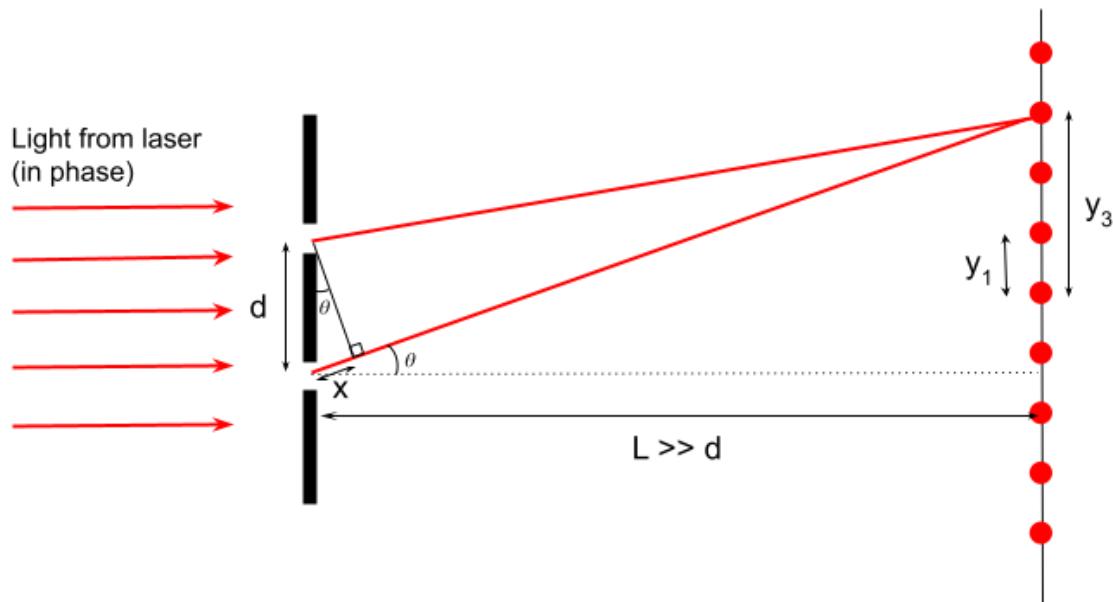


Figure 3: Mathematical Representation of Interference Pattern (not to scale)

Since the distance from the slit to the screen,  $L$ , is much greater than the slit separation distance,  $d$ , we shall assume that the two waves are parallel. From trigonometric computation, where  $\theta$  is the angle of light to the normal, we can deduce that

$$x = d \sin \theta \quad (1)$$

and

$$\tan \theta = \frac{y_n}{L}, \quad (2)$$

where  $y_n$  is the distance of the  $n^{\text{th}}$  bright spot from the center bright spot. Because  $L \gg d$ :

$$\tan \theta \approx \sin \theta, \quad (3)$$

and thus,

$$x = d \frac{y_n}{L}. \quad (4)$$

At bright spots,  $x$  is an integer multiple of the wavelength—shown by:

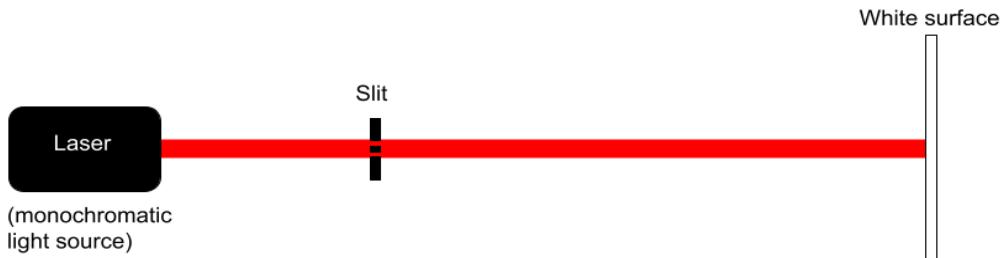
$$x = n\lambda \quad (5)$$

Combining equations (4) and (5) and then rearranging, we get a linear relationship,

$$y_n = \frac{L\lambda}{d} n \quad (6)$$

### 3. Apparatus and Procedures

Slit A (double-slit) and Slit B (40-slit), both with different separation distances between their constituent slits, were measured separately. We shone a 1.5 mW helium-neon laser, consisting of one particular wavelength, through a slit onto a white surface. The wavelength ( $\lambda$ ) of the light was 632.8 nm, and the distance from the slit to the white surface ( $L$ ) was 180.05 cm, with an uncertainty of 0.5 mm. We measured the distance,  $y_n$ , between each bright spot and the center bright spot, using a ruler with uncertainty of 0.5 mm. The separation distance between the constituent slits of Slit A and Slit B was 0.132 mm and 0.066 mm, respectively, according to manufacturer data. A diagram of our apparatus is shown in Figure 4.

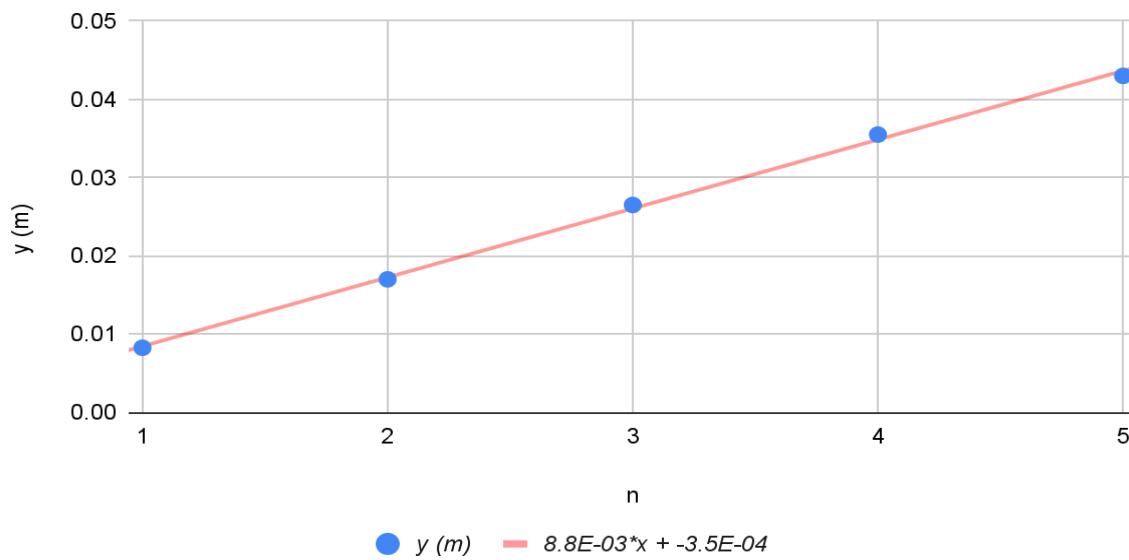


**Figure 4: Double-Slit Experiment Apparatus.**

#### 4. Data

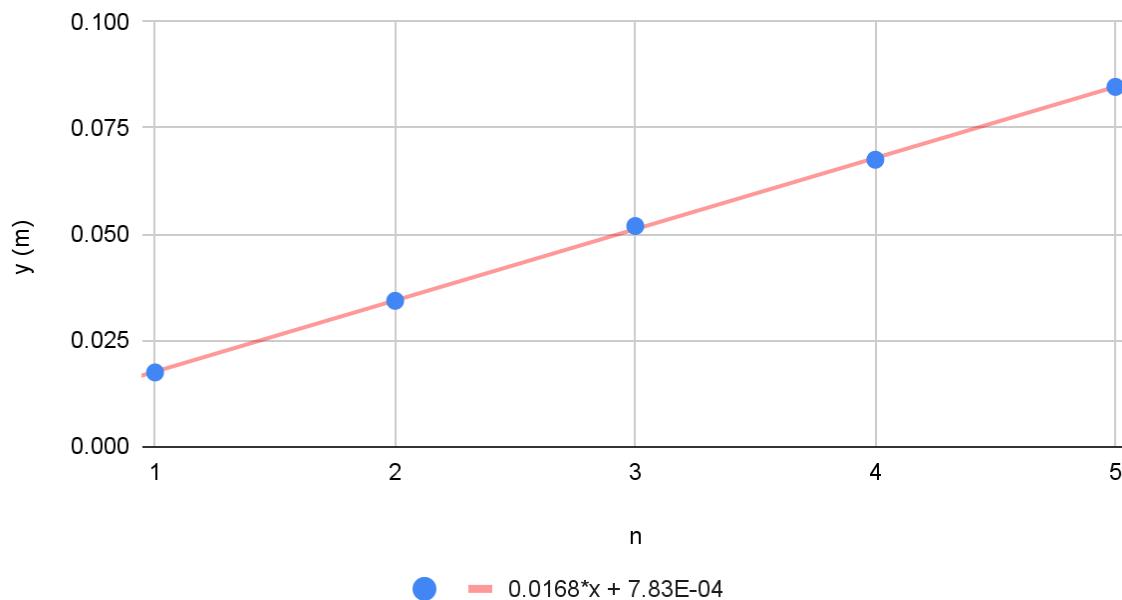
In Figures 5 and 6, we plotted  $y_n$  and  $n$  for both slots and applied a best-fit line to the data. Using the slope of the line,  $\frac{L\lambda}{d}$ , we solved for  $d$  and compared these values with the manufacturer's values, shown in Table 1. Because the manufacturer values fall within the uncertainty of the experimental values, the results are consistent with the wave theory of light.

Linear Relationship Between  $y$  and  $n$  (Slit A)



**Figure 5: Slit A: Graphic Demonstration of the Linear Relationship Between Experimental Values of  $y$  and  $n$ . Best fit line shown in red.**

## Linear Relationship Between y and n Using Slit B



**Figure 6: Slit B: Graphic Demonstration of Linear Relationship Between Experimental Values of y and n. Best fit line shown in red.**

### Example Calculation of $d$ for Slit A:

$$L = 1.8005 \text{ m}$$

$$\lambda = 632.8 \times 10^{-9} \text{ m}$$

$$\text{slope} = 0.00877 \text{ m}$$

$$\text{slope} = \frac{L\lambda}{d}$$

$$d = \frac{L\lambda}{\text{slope}}$$

$$d = \frac{1.8005 \text{ m} \times 632.8 \times 10^{-9} \text{ m}}{0.00877 \text{ m}}$$

$$d = 0.130 \text{ mm}$$

**Table 1: Experimental Values vs Manufacturer Values**

	$d$ experimental value (mm)	$d$ manufacturer value (mm)
Slit A	$0.130 \pm 0.006$	0.132
Slit B	$0.068 \pm 0.002$	0.066

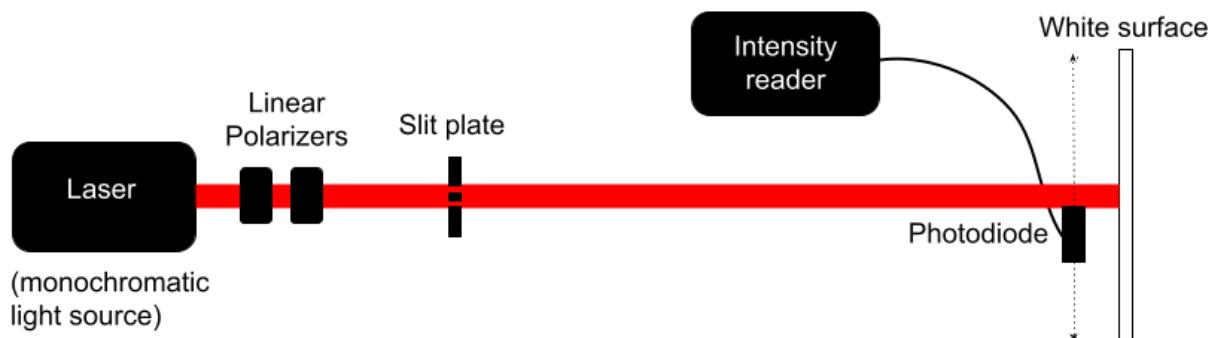
## B. Quantifying the Brightness of the Interference Pattern

Considering the peculiar nature of the experiment, we sought to confirm the existence of the interference pattern in a more reliable manner. Thus, we quantified our visual observations of the interference pattern in two ways: measuring the intensity of the light across the interference pattern and counting the number of photons across the interference pattern. We used Slit A for both experiments.

### 1. Measuring Intensity

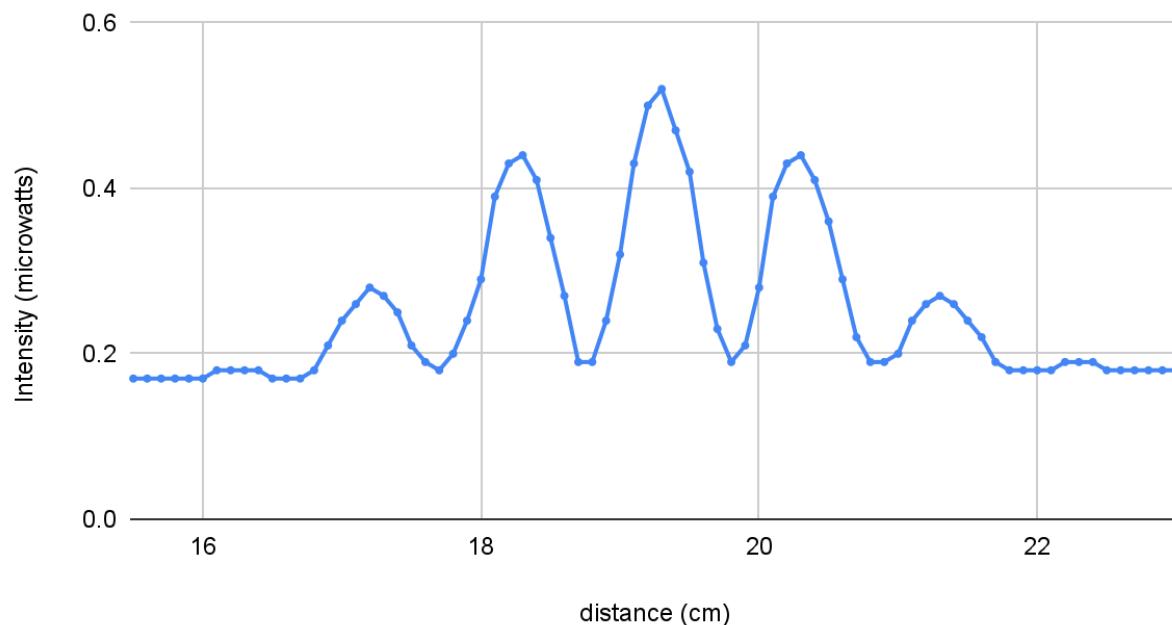
In a dark room, we shone a laser through two linear polarizers to decrease the intensity to a suitable level for the photodiode. Slit A was placed behind the polarizers and a white surface was placed behind slit A. Directly in front of the white surface, we moved a photodiode in millimeter increments across about 8 cm of the interference pattern to detect the spread of intensity. A diagram of our apparatus is shown in Figure 7.

In Figure 8, we plotted the intensity of the interference pattern against the distance. The peaks represent areas of high intensity, corresponding to bright spots, and the dips represent areas of low intensity, corresponding to dark spots. Clearly, there is an alternating pattern of bright and dark spots, and as the distance from the center bright spot increases, the intensity of the bright spots decreases. This is consistent with our previous observations.



**Figure 7: Intensity Experiment Apparatus**

### Intensity (microwatts) vs. distance (cm)



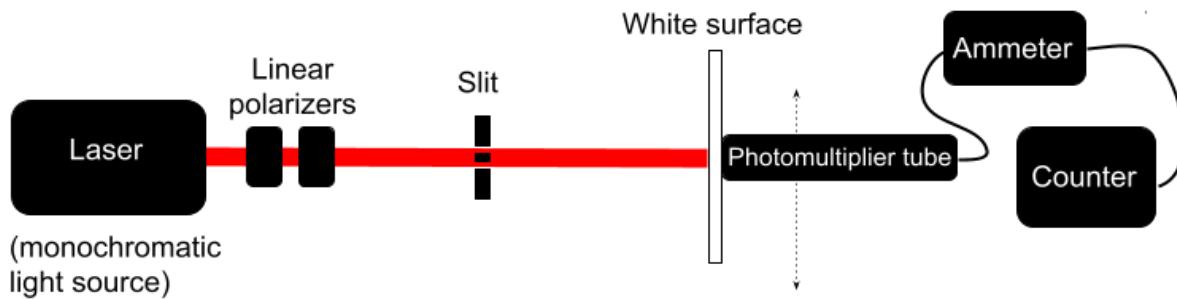
**Figure 8: Graph of Intensity across interference pattern.**

## 2. Counting photons

We now seek to investigate whether or not a similar interference pattern to that of Fig. 8 will be obtained if we reduce the intensity of the light to the point at which we can count individual photons, rather than average light intensity. Using a similar setup, we shone a laser through two linear polarizers to reduce the intensity of the light to suit the limits of the apparatus. Slit A was placed behind the linear polarizers and a white surface was placed behind Slit A. A photomultiplier tube was shifted horizontally in millimeter increments directly behind the screen. The photomultiplier tube amplifies the signal of each photon so that the counter is able to detect the photons. A diagram of the apparatus is shown in Figure 9.

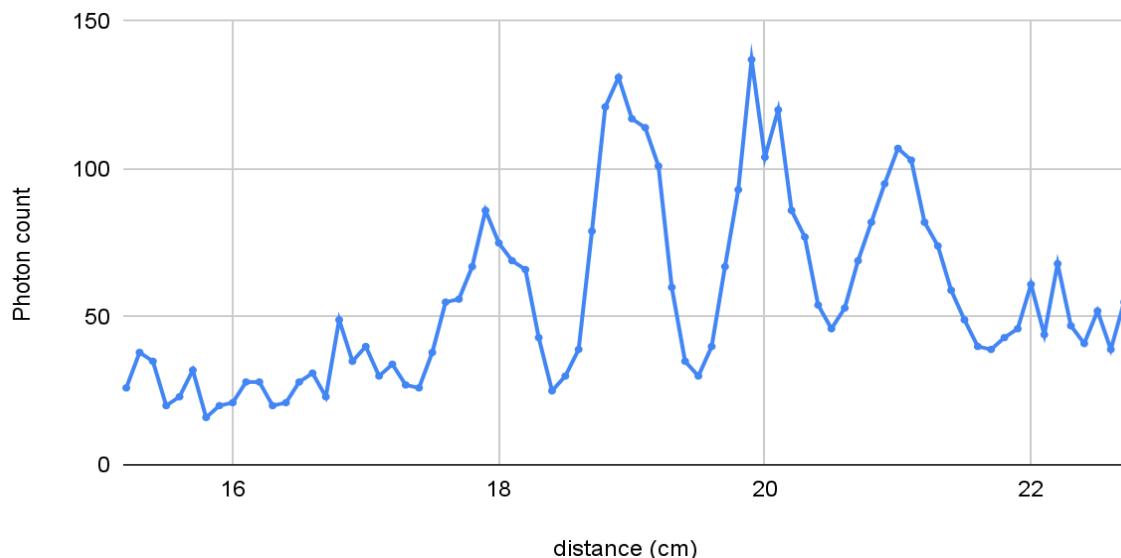
To count the number of photons across the interference pattern, we placed the photomultiplier tube at a distance of 15.2 cm. The number of photons at this distance was counted for 30 seconds. Then, we moved the photomultiplier tube one millimeter and counted for another 30 seconds. This method was repeated until a range of approximately 8 cm was measured.

In Figure 10, we plotted the photon count against the distance. The peaks in photon count correspond to bright spots, and the dips correspond to dark spots. The data is consistent with the observed pattern produced in the continuous wave experiment (Fig. 7) – there is an arrangement of alternating dark and bright spots, and the photon count at bright spots decreases as they get further away from the center bright spot. Based on these two experiments, the interference pattern we observed is essentially the same. Both the continuous-wave model and the single-photon model yield similar results.



**Figure 9: Photon Counting Apparatus**

Photon count vs. distance (cm)

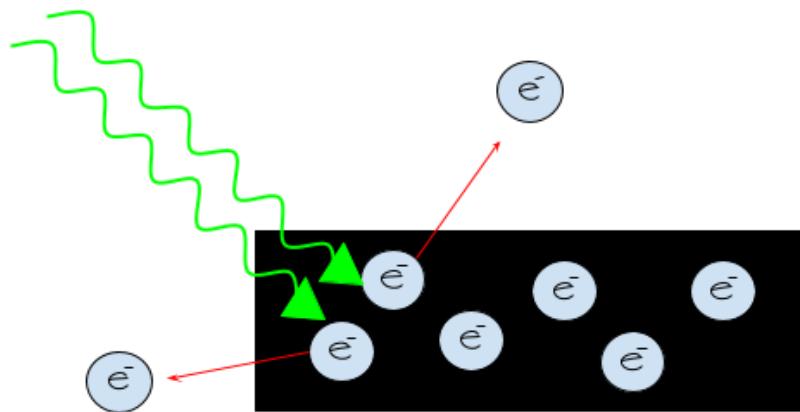


**Figure 10: Graph of Photon Counts Across Distance.**

## C. Photoelectric Effect

### 1. Theory

When light strikes a metal, electrons are released from its surface—this is called the photoelectric effect (Figure 11). In accordance with the classical wave theory of light, physicists predicted that increasing the frequency of the incoming light would not change the kinetic energy of the liberated electrons. Whilst their hypotheses were intuitive, experimentation demonstrated the contrary: increasing the frequency of the light increased the kinetic energy of the electrons.<sup>8</sup> Describing light as an electromagnetic wave proved an inadequate approximation of light's nature.



**Figure 11: Photoelectric Effect**

In 1905, Albert Einstein created a new theory to explain the effect. Echoing Newton's corpuscular theory of light, Einstein proposed that light could be described in terms of particles called photons, which had a specific energy,  $E$ , depending on the light's frequency. Specifically, he theorized that  $E = hf$  where  $h$  is Plank's constant and  $f$  is the frequency of the light. When photons collide with electrons in the metal, they transfer all their energy to the electron. The electron uses a certain amount of this energy,  $\Phi$ , to escape the forces of the metal.  $\Phi$  is the work function of the specific type of metal—the amount of energy needed to remove an electron from the metal's surface. The remaining energy is transformed into kinetic energy. We can write this as:  $hf = \Phi + K$ , where  $K$  is the kinetic energy. Importantly, if the photon has less energy than the work function,  $\Phi$ , electrons will not be emitted from the metal. Photons must meet a minimum threshold frequency,  $f_0$ , to meet the work function:  $\Phi = hf_0$ . So,

$$hf = hf_0 + K \quad (7)$$

Equation (7) shows that the kinetic energy of the electrons are directly proportional to the frequency of the incoming light, since the work function is constant. Experimentally, in order to measure the kinetic energy of the emitted electrons, we can use the stopping potential. If a negative potential difference is applied to the anode, against the movement of the electrons, the current will decrease because the opposing force of the voltage increases the kinetic energy electrons need to cross the gap. Using this property, we determine how much voltage is required to stop the current of electrons. This “stopping potential”,  $V_s$ , is directly related to the kinetic energy of the electrons:

$$K = V_s e \quad (8)$$

where  $e$  is the charge of an electron  $1.602 \times 10^{-19} C$ . Substituting equation (8) into (7) yields,

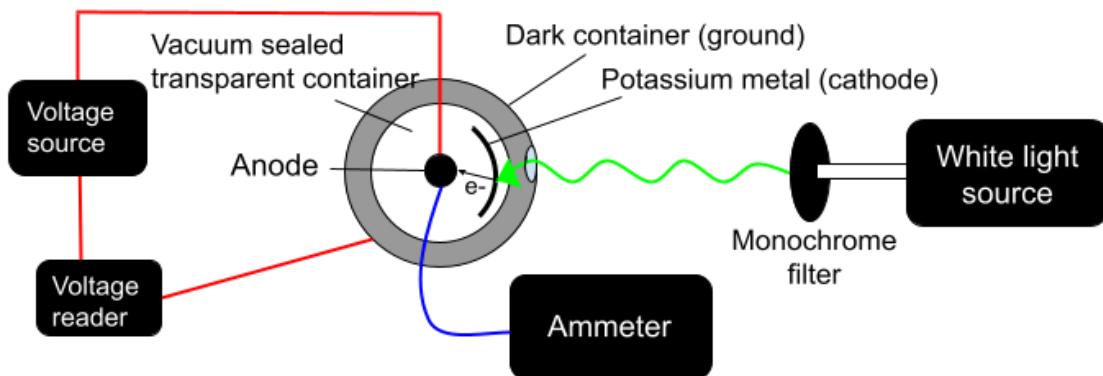
$$hf = hf_0 + V_s e \quad (9)$$

Rearranging (9), we get a linear relationship:

$$f = \frac{e}{h} V_s + f_0 \quad (10)$$

From equation (10), we can deduce that as the frequency of the incoming light increases, the magnitude of the stopping potential also increases.<sup>9</sup>

## 2. Apparatus and Procedure



**Figure 12: Photoelectric effect apparatus**

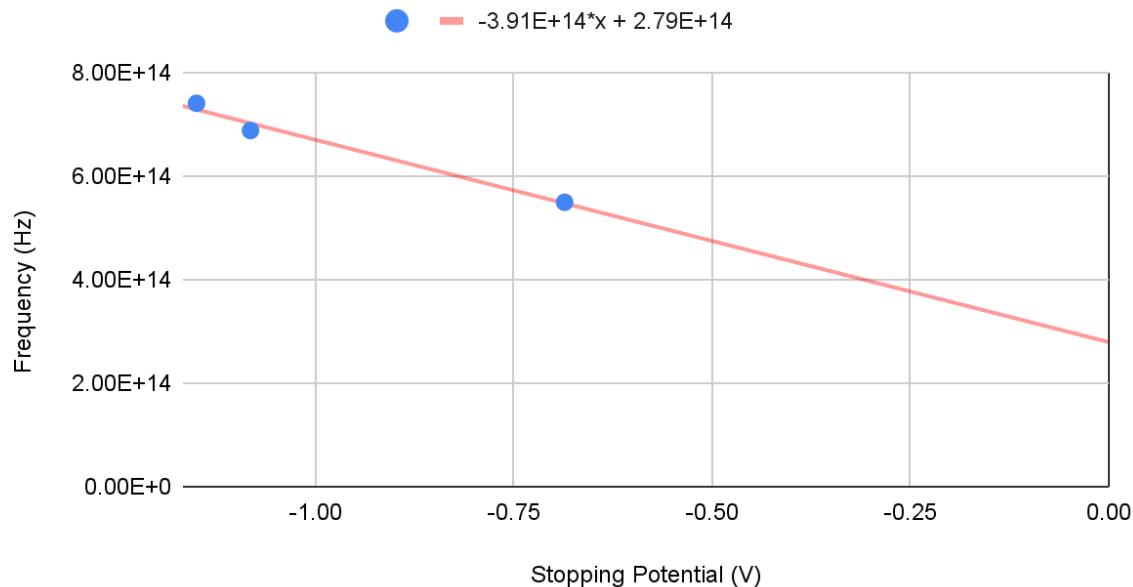
In a dark room, we shone a white light consisting of several different wavelengths through each of three monochrome filters of different wavelengths: 546, 436, and 405 nm. The uncertainties for these filters were 10 nm each. The filtered light then traveled into a closed container with a transparent vacuum-sealed compartment containing a potassium metal cathode and a grounded anode. The anode is connected to an adjustable voltage source and an ammeter. A diagram of the apparatus is shown in Figure 12.

To measure the stopping potentials for each wavelength of light, we adjusted the voltage until the ammeter detected no current. We recorded this voltage as the stopping potential. For each of the three wavelength filters, we made three measurements.

## 3. Data and Analysis

In Figure 13, we plotted the frequency of light against the measured stopping potentials and applied a best fit line. As Einstein's particle theory predicted, the magnitude of the stopping potential increased with increasing frequency of light. Therefore, our results are consistent with Einstein's particle theory of light.

## Linear Relationship Between Frequency and Stopping Potential



**Figure 13: Graph Showing Linear Relationship Between Frequency of the Light and Stopping Potential.**

## III. Conclusions

Taking light as a wave, we experimentally determined the slit separation distances for both the Double-Slit A and Forty-Slit B apparatus by measuring the resulting laser interference patterns. The resulting experimental values were consistent, within the margins of uncertainty, with the manufacturer-specified values of the slit separations. Therefore, we concluded that the wave theory of light provides an adequate explanation of the interference pattern shown in the Double-Slit experiment. We then reduced the intensity of the incident light to the point of counting single photons and repeated the measurements of the Forty-Slit apparatus, using a photomultiplier tube detector. The resulting interference pattern was similar to that obtained in the continuous-wave limit, using a photodiode.

We also examined the photoelectric effect, which the wave theory is unable to explain. We found that Einstein's equations align with our experimental data. Hence, we also accept the particle theory of light.

Because both theories alone are inadequate in explaining all phenomena, we postulate that light expresses either wave or particle characteristics, depending on its interactions with particular circumstances such as the apparatus. However, we suspect that there is a fundamental idea that can explain both the wave and particle natures of light.

## IV. Future Research

For future experiments, we would like to venture beyond the classical world with its particle and wave theories to explore light quantum mechanically. Because the underlying unifier of light's two natures is unknown, tinkering would be our next step, so we would utilize the wave function—a quantum mechanical function that tells us the properties of particles in stochastic terms—to ascertain what we can do if we were

to treat light's nature as variable from the onset. And, since we would be dealing with probabilities in the quantum mechanical realm, we would also attempt to count the number of photons before they pass through the slits and see if that aligns with the count of photons in the interference pattern on the screen. For the photoelectric effect, we would also like to add more data points and use different metals to see if that improved agreement with Einstein's particle theory.

## V. Acknowledgements

We thank Dr. Barry Luokkala for his tremendous efforts in assembling suitable apparatus for our project, for his helpful advice and guidance, and for his decades-long efforts in preserving the message of PGSS—bringing together young scientists from around Pennsylvania and allowing them to thrive. We are grateful to our Faculty advisor, Beka Modrekiladze, for allotting his time and expertise to aiding our research and explaining difficult concepts. We appreciate TA extraordinaire Haley Williams for their many hours of dedication to assisting in our research, and providing us with both intellectual and emotional support throughout the program.

Likewise, we are also thankful to Carnegie Mellon University, PGSS Corporate Sponsors, the PGSS Alumni Association and all other PGSS donors, as without their support, our scientific curiosities would not have been granted the expertise and resources needed to flourish.

## VI. References

<sup>1</sup> Olivier Darrigol, *A History of Optics from Greek Antiquity to the Nineteenth Century*, (Oxford University Press, Oxford, 2012), 1-3, 17-18, 66

<sup>2</sup> Isaac Newton, "A letter of Mr. Isaac Newton, professor of the mathematicks in the University of Cambridge; containing his new theory about light and colors," *Philosophical Transactions of the Royal Society* **6**, (80), 3075-3087, (19 February 1671/2)

<sup>3</sup> Patricia Fara, "Newton shows the light: a commentary on Newton (1672) 'A letter ... containing his new theory about light and colours...,'" *Philosophical Transactions of the Royal Society* **373**, (2039), (13 April 2015)

<sup>4</sup> Arvind Kumar, "Newton's contributions to optics," *Resonance* **11**, 10-20, (December 2006).

<sup>5</sup> Kenneth R. Spring and Michael W. Davidson, "Light: particle or a wave?," *Florida State University*, Accessed August 3, 2023, <https://micro.magnet.fsu.edu/primer/lightandcolor/particleorwave.html>

<sup>6</sup> "Photoelectric effect," *Britannica Encyclopedia*, (7th July 2023),  
<https://www.britannica.com/science/photoelectric-effect>

<sup>7</sup> Paul Peter Urone and Roger Hinrichs, "College physics," *OpenStax*, chapter 27.4, (21 June 2012),  
<https://iu.pressbooks.pubopenstaxcollegephysics/chapter/youngs-double-slit-experiment/>

<sup>8</sup> "Photoelectric effect," *Khan Academy*, Accessed August 3, 2023,  
<https://www.khanacademy.org/science/physics/quantum-physics/photons/a/photoelectric-effect>

<sup>9</sup> "The photoelectric effect and Planck's constant," *Saint Monica College*, Accessed August 3, 2023,  
[https://www.smc.edu/academics/academic-departments/physical-sciences/physics/documents/physics-23/Phys\\_23\\_T12\\_The\\_photoelectric\\_effect\\_and\\_Plank's\\_constant.pdf](https://www.smc.edu/academics/academic-departments/physical-sciences/physics/documents/physics-23/Phys_23_T12_The_photoelectric_effect_and_Plank's_constant.pdf)

# Crystal Structure and Differences in Magnetic Behavior in Copper Sulfate and its Hydrates

Carter Brown, Brighton Risch, Ronald Wang

## Abstract

The magnetic properties of materials have long fascinated researchers due to the wide range of applications that magnetic materials have on technology. This study aims to unravel the relationship between crystal structures and their magnetic behaviors by delving into the magnetic susceptibility of copper sulfate and its hydrates. In total, we compared three variations of copper sulfate: anhydrous, monohydrate, and pentahydrate. To probe the magnetic properties and crystal structure, we performed electron spin resonance, x-ray powder diffraction, and magnetic susceptibility measurements. Our findings revealed distinct variations in magnetic susceptibility among the various hydrates of copper sulfate, suggesting that the presence and arrangement of water molecules within the crystal lattice significantly influence their magnetic characteristics.

## I. Introduction

### A. Background

Copper(II) sulfate is an inorganic compound that forms hydrates from monohydrate to heptahydrate. As the number of water molecules within copper(II) sulfate increases, the crystal structure is distorted and compressed. The resulting hydrate has significantly different material properties associated with this structural change. Despite the importance of understanding the origins of the magnetic properties of materials, little research has been conducted on the correlation between the different crystal structures of inorganic hydrates and its effects on their magnetic behaviors. Additionally, the latest research done looking into the magnetic susceptibility and magnetic behavior of the various copper(II) sulfate hydrates is from the 1960s<sup>1</sup>. Anhydrous, monohydrate, and pentahydrate copper(II) sulfates, shown in Figure 1, are all known to be paramagnetic based on measurements of their magnetic susceptibility; however, how these magnetic susceptibilities are influenced by the crystal structures of these different compounds is not readily known. In order to better identify such differences within anhydrous copper(II) sulfate and its hydrates, x-ray diffraction was used to identify and compare crystal structures. Electron spin resonance spectroscopy was then done in order to determine the individual interactions of electrons with an external magnetic field. Lastly, the measurement of the magnetic susceptibility was made with a Gouy balance.

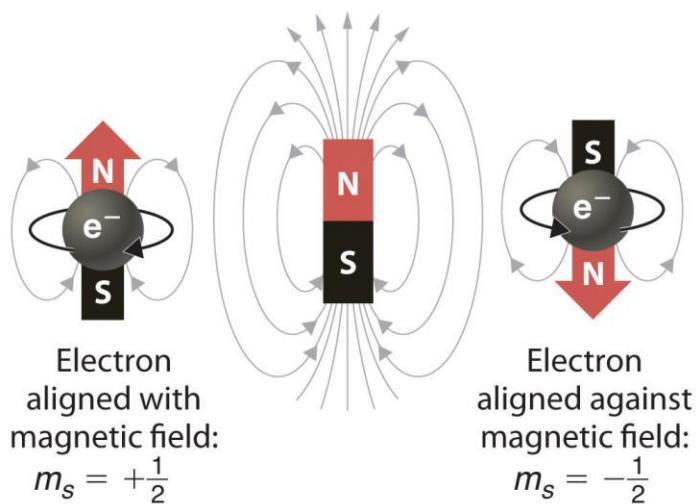


**Figure 1: Samples of anhydrous, monohydrate, and pentahydrate copper(II) sulfate respectively<sup>2</sup>**

## II. Theory

### A. Electron Spin Resonance Spectroscopy

Electron spin resonance (ESR) is a spectroscopic technique used to study the behavior of unpaired electrons in materials. In quantum mechanics, electrons possess an intrinsic property called spin, which is a form of angular momentum. Electrons are fermions, which means they take on half-integer values for spin. An electron can either be  $+1/2$  (spin-up) or  $-1/2$  (spin-down). When an electron is placed in an external magnetic field, the energy of the electron splits into a lower energy state that is aligned with the field, and a higher energy state that is unaligned. The energy separation between the spin states is directly proportional to the strength of the applied field. This effect is called the Zeeman effect, and the energy splitting of the two spins is named Zeeman splitting.



**Figure 2: Electrons in an external magnetic field<sup>3</sup>**

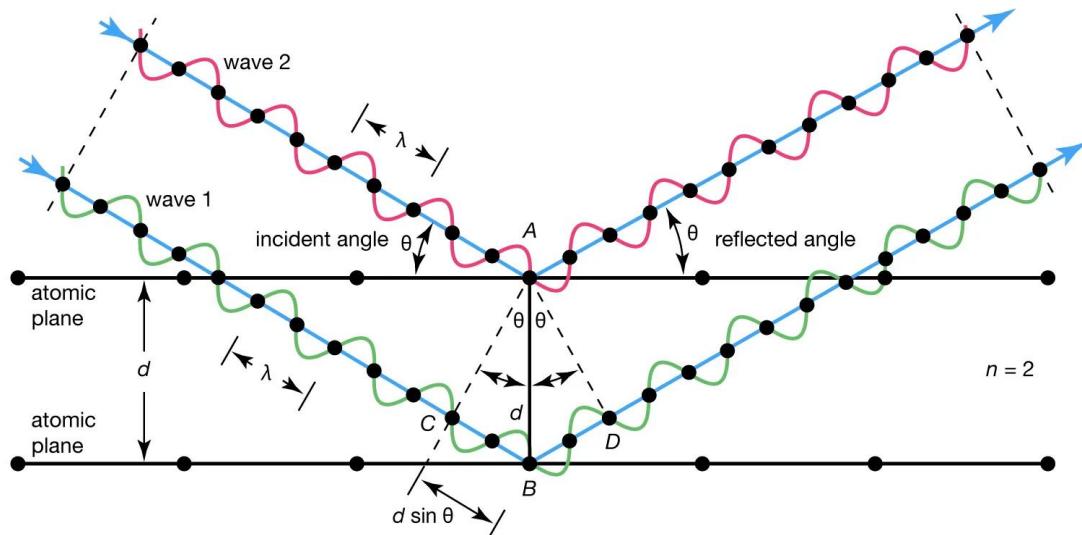
Magnetic resonance occurs when a system absorbs electromagnetic energy at a specific resonance frequency that matches the energy difference between the two spin states. This corresponds to the Zeeman splitting between the two possible spin states of the unpaired electron in the external magnetic field. In this example, the electron on the left of Figure 2 is in a lower energy state than the electron on the right because it is aligned with the magnetic field. ESR experiments show how the electrons in a material interact with microwave radiation when put in a magnetic field.

### B. X-ray Diffraction

X-ray powder diffraction (XRD) is a powerful analytical technique used to study the structure of crystalline materials by providing valuable information about the arrangement of atoms or molecules within a substance.

X-rays are a form of electromagnetic radiation with very short wavelengths, typically on the order of 10-10 meters. When x-rays interact with a crystal, they scatter off the regularly spaced atomic planes within the crystal structure. Depending on the structure of the crystal and the angle of the incident beam, the x-rays may form a constructive or destructive diffraction pattern. XRD graphs are then created by measuring what

angles constructive interference occurs for. This is important because constructive interference only occurs at specific angles and between specific layers, giving each crystal structure a unique diffraction pattern. From this pattern, researchers are able to deduce information about the crystal structure.



**Figure 3: Example of constructive interference between two atomic planes.<sup>4</sup>**

### C. Magnetic Susceptibility

Magnetization is the measure of the density of both permanent and induced dipole moments in a material. The magnetic susceptibility is a factor that will affect the magnetization of a material when in an external applied magnetic field. The extent of magnetization is defined by the equation:

$$\mathbf{M} = \chi \mathbf{H}$$

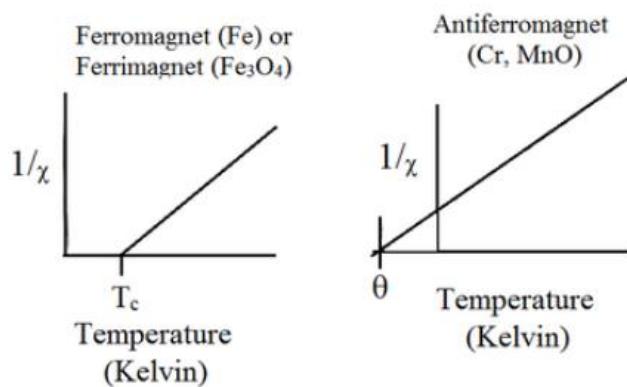
Where M is the magnetization of a material,  $\chi$  is the magnetic susceptibility, and H is the strength of an applied magnetic field.

The magnetic susceptibility is a temperature-dependent quantity. For ferromagnetic materials,  $\chi$  follows the equation:

$$\chi = \frac{C}{T - \theta_{CW}}$$

Where C is the Curie constant for any given material, T is the temperature, and  $\theta_{CW}$  is the Curie-Weiss temperature. Depending on the sign and value of the Curie-Weiss temperature, the material can be classified as paramagnetic, ferromagnetic, or antiferromagnetic. If  $\theta_{CW}$  is 0, the material is paramagnetic.

Paramagnetic materials, such as aluminum and gold, are weakly attracted to magnetic fields. However, as the temperature approaches zero, the magnetic susceptibility diverges to infinity. In ferromagnetic materials,  $\theta_{CW}$  is positive so the magnetic susceptibility diverges at a temperature before zero. Above this temperature, the material will be weakly magnetic; however, below this temperature, the material is extremely magnetic. Conversely, antiferromagnetic materials have a  $\theta_{CW}$  that is negative. This means that the magnetic susceptibility will diverge at a negative temperature, which is not physically possible. This means that even as the temperature approaches 0°K, the magnetic susceptibility will not infinitely diverge. In fact, as the temperature approaches zero, the inter-particle interaction within the material causes neighboring magnetic moments to alternately align and anti-align with the external field, giving a net magnetization of zero.



**Figure 4: Graph of  $1/\chi$  vs temperature for ferromagnetic and antiferromagnetic materials.  $T_c$  is another way to express the Curie-Weiss constant for ferromagnetic materials.<sup>5</sup>**

### III. Method/procedure

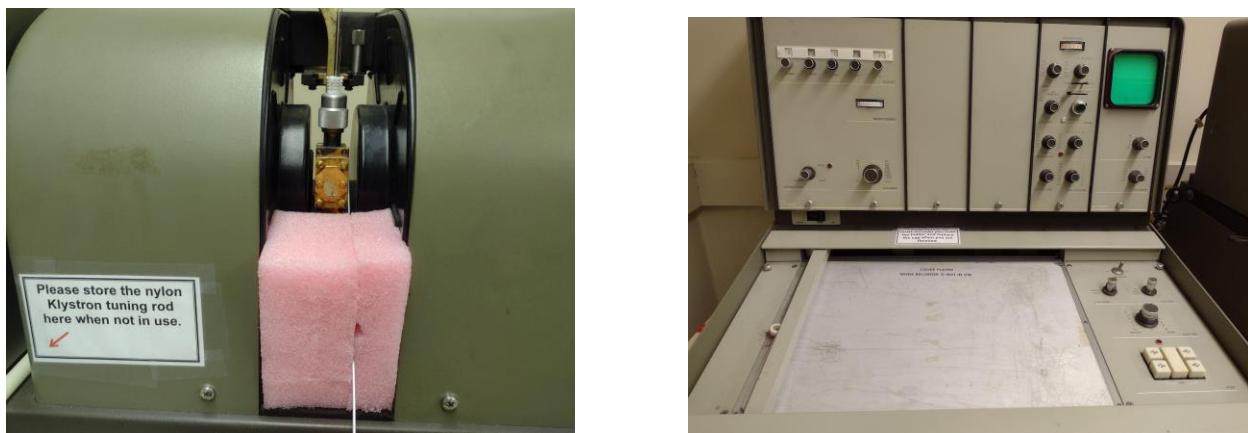
#### A. Electron Spin Resonance Spectroscopy

The anhydrous copper sulfate sample was already prepared, so only the monohydrate and pentahydrate samples had to be prepared by grinding the crystals using a mortar and pestle, then putting the powdered crystals into a pyrex tube. We used a DPPH sample to calibrate the machine and to make sure it was returning accurate results. The ESR machine is depicted in figure 4, 5. Figure 4 shows the electromagnets that generated the magnetic field and the sample which is held in between the two. Figure 5 shows the part of the machine that drew the results of the machine on paper. We did not use this because we took final measurements on an electric oscilloscope. This machine had a consistent microwave frequency and searched over a range of magnetic field strengths to find peaks. We initially started searching over a broad range of field strengths (2000 gauss). Once we found the center of the peaks, we reduced the field range depending on the material. Table 1 lists the settings which were adjusted to find the peaks. The most important settings are gain, field set, and scan range. Gain multiplied the data so visually, there were larger peaks that we could identify. Field set was the zero point which was the middle of the measurement. We wanted to get the field set as close to the middle of the peak as possible. Finally, the scan range adjusted the range of magnetic field strengths the machine scanned over. A smaller range means larger peaks and more precise data. The data was accepted when there was as little random noise as possible, it clearly showed peaks, and it was close to expected results if we had anything to compare it to.

## ESR Machine Settings

**Table 1: Settings of ESR machine**

Field set	Moves the data horizontally
Output zero	Moves the data vertically, always start in the middle
Gain	Stretches the data vertically
Scan range	Stretches the data horizontally
Time constant	Changes the response time to changes in the measurement
Scan time	How long the scan runs for
Power	Increases or decreases power of microwaves
Frequency	Increases or decreases microwave frequency
Mode	Standby, tune, operate



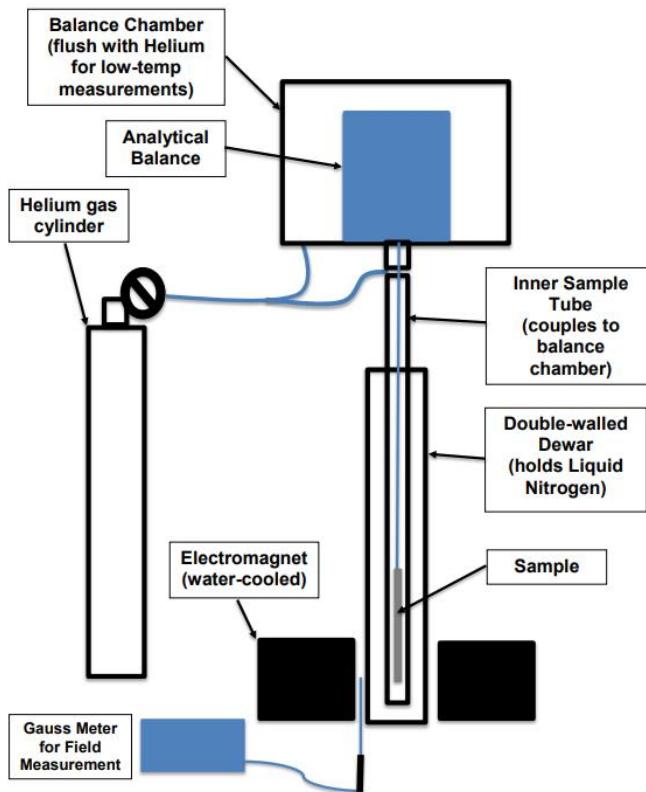
**Figure 5: ESR machine**

## B. X-Ray Diffraction

XRD was carried out on a Siemens D500 X-ray Diffractometer with Cu target x-rays ( $1.540598 \text{ \AA}$ ) on all three samples. After powdered samples were loaded into the diffractometer, the shutter was opened and the x-ray beam was allowed to hit the target sample at angles of  $2\theta = 20$  to  $70$  degrees, as calibrated by the goniometer. A detector collected the diffracted light from the x-rays off of the copper(II) sulfate sample and transmitted via MDI Databox to the MDI DataScan software. The obtained diffraction pattern was compared to a database of known crystal structures to identify the material through phase identification. The software used for phase identification was Profex.

### C. Magnetic Susceptibility

We used a Gouy balance to measure the extent of magnetization. A Gouy balance, shown in Figure 6, contains a balance at the top, which is attached to the sample via a long wire. The sample is lowered into a strong magnetic field and the difference in mass between no magnetic field and an external magnetic field is recorded.



**Figure 6: Diagram of the Gouy Balance Apparatus<sup>6</sup>**

We prepared samples of anhydrous, monohydrate, and pentahydrate copper sulfates in Pyrex tubes, shown in Figure 7. We did the room temperature analysis for all three samples first. The sample was loaded into the liquid nitrogen container to block air circulation. Data for each sample was collected five times each at 3 A, 6 A, 9 A, and 12.5 A current for each sample. Tests under liquid nitrogen were done next to conduct low-temperature measurements at 77K. We loaded in the sample, then displaced the air in the scale and the liquid nitrogen container with helium. The liquid nitrogen was poured in until the entire sample was surrounded by liquid nitrogen. More liquid nitrogen was added as needed throughout the experiment. After waiting for the sample to completely cool, we took the data the same way as room temperature.



Figure 7: Pentahydrate, monohydrate, and anhydrous copper(II) sulfate samples

## IV. Results/Analysis

### A. Electron Spin Resonance Spectroscopy

The ESR spectrometer generated a graph of the spectra of the copper sulfates. This ESR spectra is the derivative of the absorption spectrum. In order to see the absorption spectrum, the trapezoidal rule was used in order to approximate the integral of the raw data. The zeroes of the ESR spectra generated by the spectrometer were used in order to determine the location of the peaks of the absorption spectrum. The ESR spectra data of DPPH as an example of the standard g factor. The g factor for a free electron is 2.0023193<sup>4</sup>. As DPPH is a free radical, its g factor should be approximate to such a value. This allows DPPH to be used as a calibration standard for ESR. The g-factor is calculated through an equation relating the absorption of energy to the magnetic field, shown below, in which  $\mu_B$  represents the Bohr magneton,  $g_e$  is the g-factor, and  $B_0$  is the external magnetic field<sup>7</sup>. The magnetic field was calculated through knowing the scan range of the spectra and the total time at which the spectra were done. The g-factor characterizes the magnetic moment of the material. The main use of measuring the g factor in this research is to qualitatively compare them to the standard DPPH g-factor to observe which differences =additional molecules might have on anhydrous copper(II) sulfate's magnetic properties.

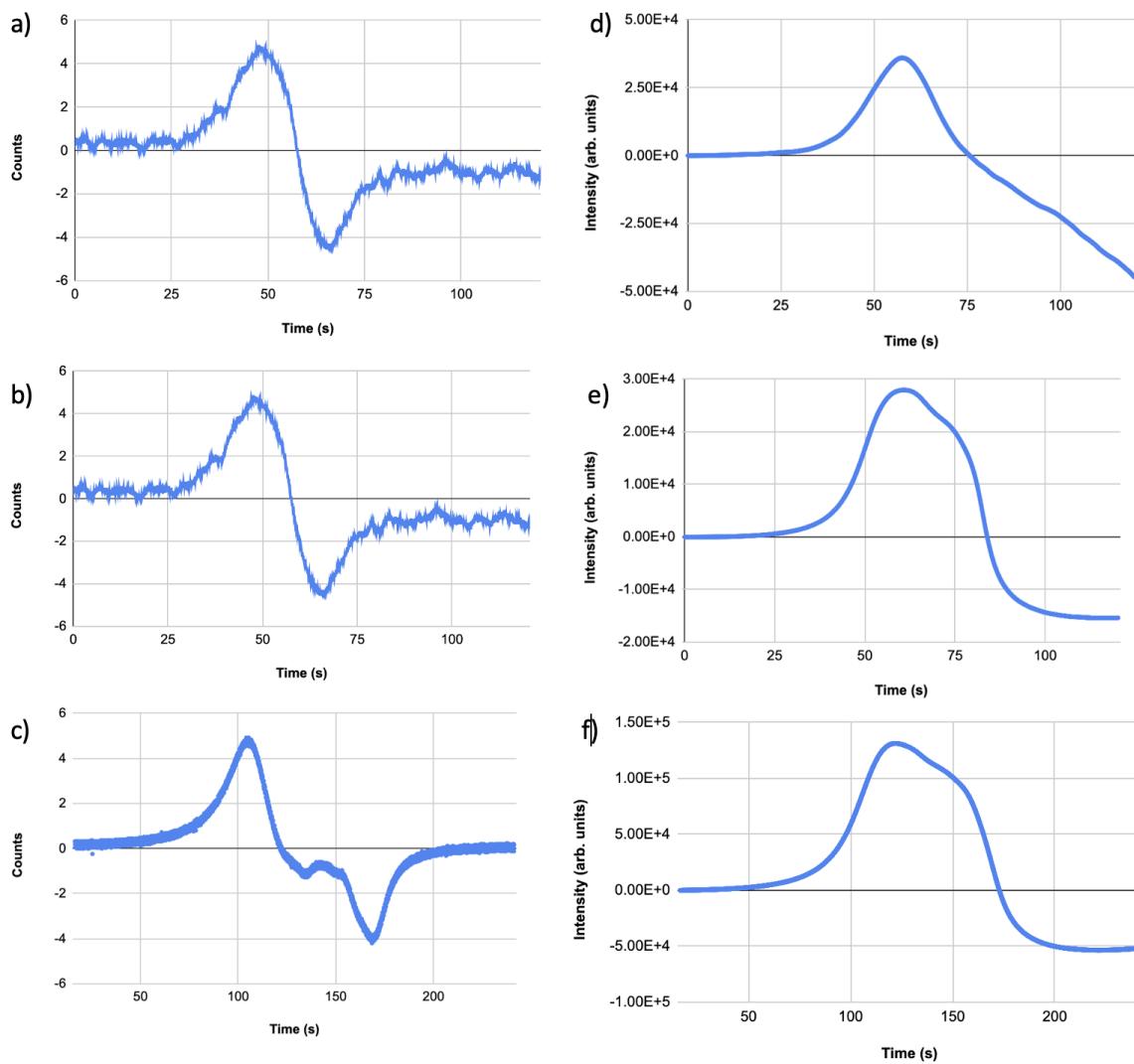
$$\Delta E = h\nu = g_e \mu_B B_0$$

**Table 2: g-factors of anhydrous, monohydrate, and pentahydrate Copper(II) Sulfates**

Crystal	Measured g <sub>1</sub> -factor	Measured g <sub>2</sub> -factor
Anhydrous CuSO <sub>4</sub>	2.168000 ± 0.000010	N/A
Monohydrate CuSO <sub>4</sub>	2.201200 ± 0.000012	2.138450 ± 0.000012
Pentahydrate CuSO <sub>4</sub>	2.199000 ± 0.000012	2.104917 ± 0.000012
DPPH	2.017099 ± 0.000010	N/A

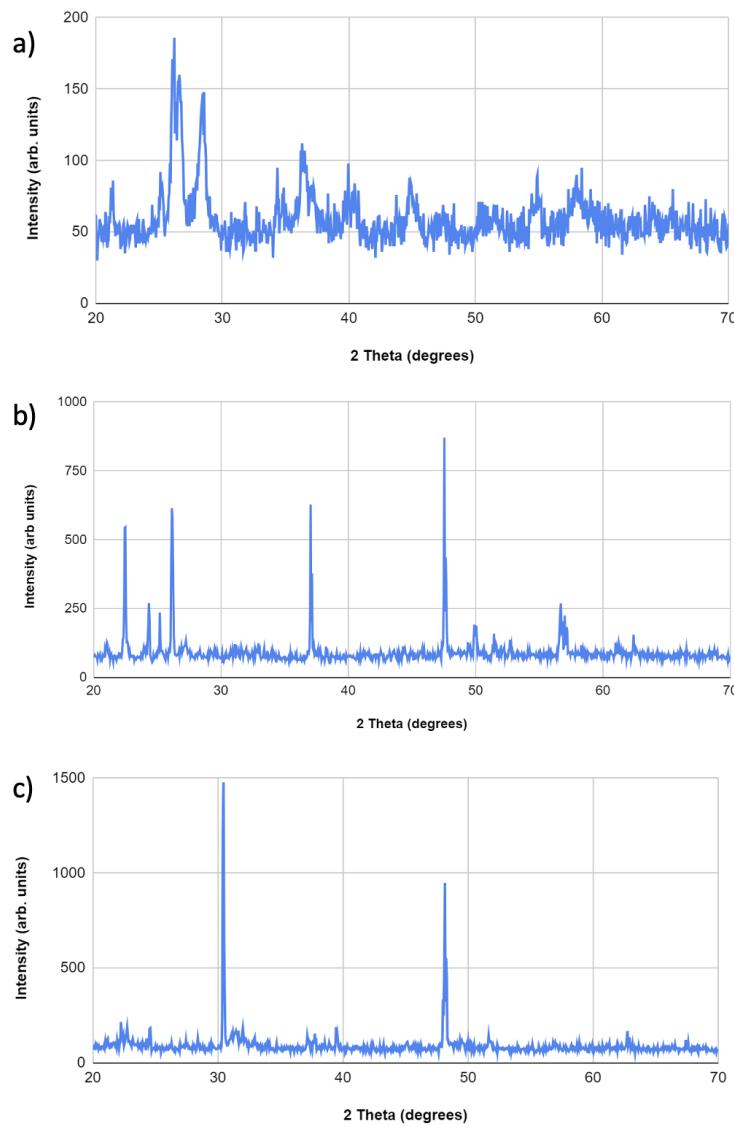
As seen in Table 2, the g-factors of the monohydrate and pentahydrate are slightly different from that of the anhydrous copper(II) sulfate. This slight change in magnetic moment may indicate that when water molecules are added to copper(II) sulfate, the crystal's magnetic properties change slightly. This change may be due to structural changes, which we explore later in this paper. Furthermore, unlike the anhydrous compound and DPPH, the monohydrate and pentahydrate crystals had an additional smaller peak in their absorption spectrum as seen in Figure 9 and Figure 10. The g-factors of these peaks were also calculated and are labeled as g<sub>2</sub> factors in Table 2.

The ESR spectra, also known as the derivative of the absorption spectrum, as seen in Figure 8 a, b, c, allow us to find points of 0 magnetic field intensity, as well as calculate the specific g-factor. To better visualize the free electron peaks of the material, we integrated the ESR derivative spectra of our monohydrate, pentahydrate, and anhydrous Copper(II) sulfate, resulting in the absorption spectra. This was done through using the trapezoidal rule in Microsoft Excel. Figures 8 d, e, f show the resulting integrated ESR derivative spectra which represent their respective absorption spectra. An interesting result of plotting such graphs is seeing the extra peak or bump the absorption spectrum of the hydrates. Furthermore, the second peak of the pentahydrate spectrum is slightly flatter than that of the second peak of the monohydrate spectrum. This difference can be attributed to the difference in the g-factor of the second peaks in Table 1. General comparisons show that each copper(II) sulfate has different ESR derivative and absorption spectra, which was an interesting discovery due to the assumption that water must have little to no impact on the ESR spectra, since water has all paired electrons and no free electrons. Such differences suggest that the addition of the water molecules must indirectly affect the ESR spectroscopy. We approach how the addition of water molecules may affect the materials' crystal structure and magnetic properties in our X-ray diffraction experiment.



**Figure 8: Derivative spectra (a, b, c) and absorption spectra (d, e, f) of anhydrous, monohydrate, and pentahydrate copper(II) sulfate**

## B. X-ray Diffraction



**Figure 9: XRD results of anhydrous (a), monohydrate (b), and pentahydrate (c) copper(II) sulfate**

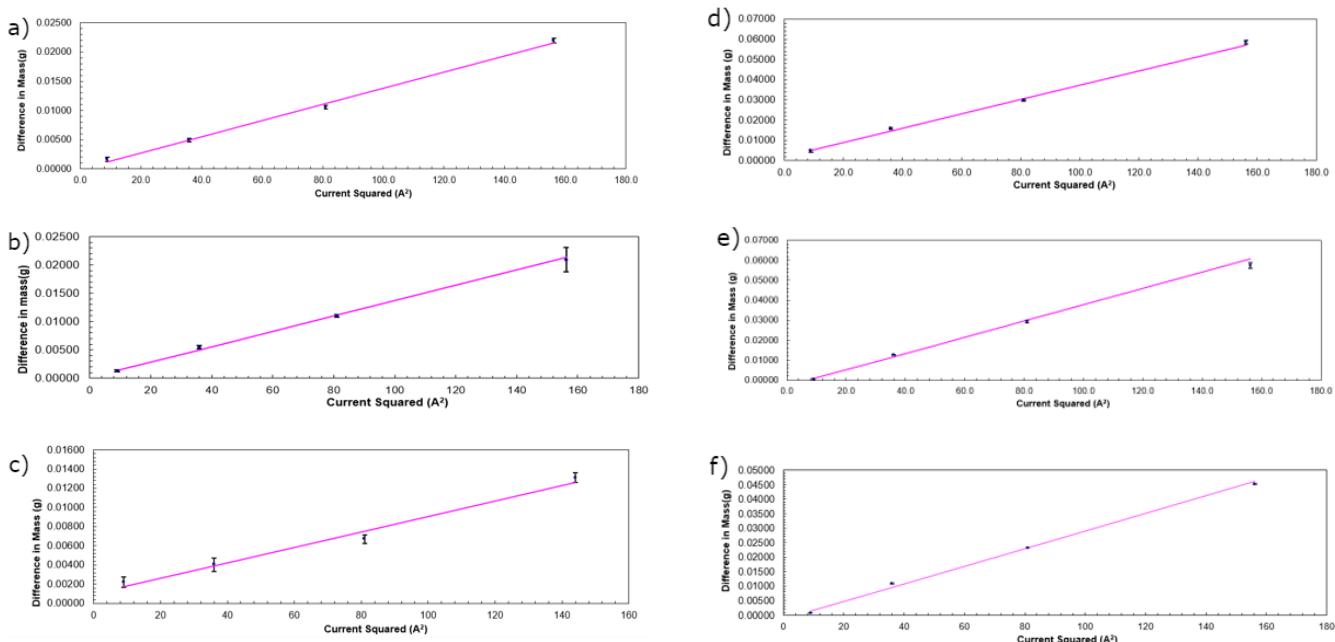
X-ray diffraction was used to confirm that the crystal structures of separate hydrates were different, as well as to confirm that the crystal structure matched with what was expected for each compound. We observe that each hydrate's crystal structure diffracts light differently, resulting in different peaks in the XRD spectra. It is therefore evident that the crystal structure for each compound is different. When we used Profex for phase identification, the x-ray diffraction pattern for each compound closely matched with existing literature values, and Profex determined that it was very likely that the copper(II) sulfate compounds matched their identity. This was a particularly interesting result because it indicates that the copper sulfate derivatives had distinct crystal structures which may have contributed to differences in the ESR spectrum. Profex is also able to determine if multiple crystal structures are present, and in this case, it determined that there was only one crystal for each of the samples. This is important because it rules out the possibility that our sample was contaminated.

### C. Magnetic Susceptibility

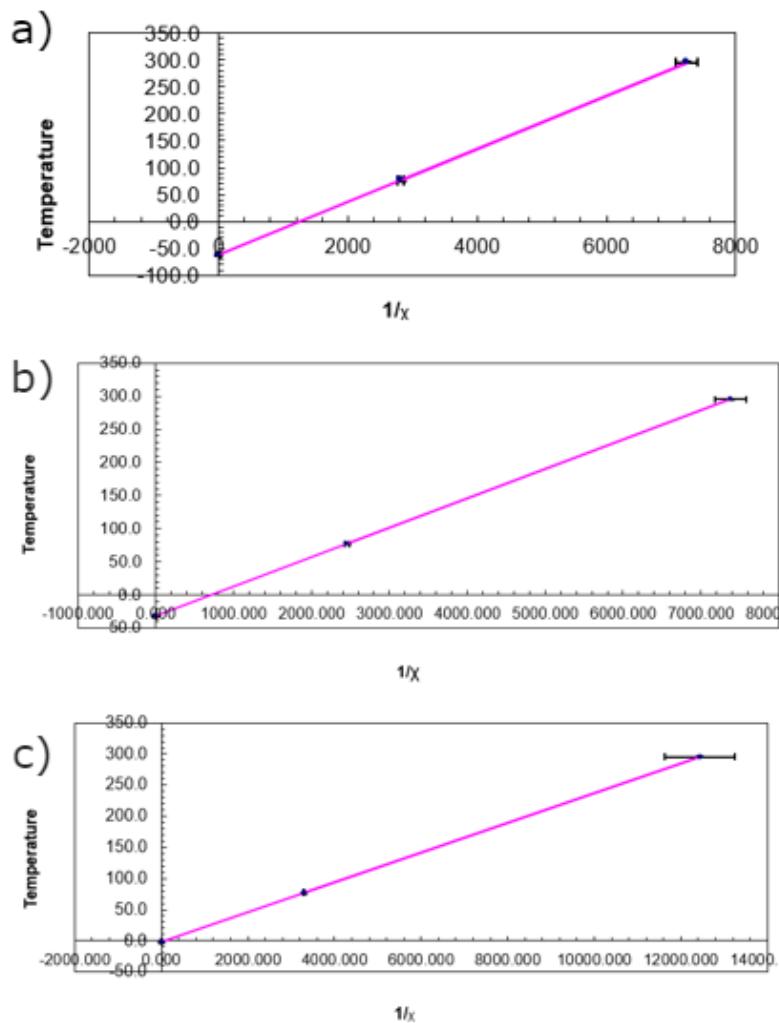
The method used in measuring the magnetic susceptibility of the samples of copper(II) sulfates was the Gouy balance method. From the equation below, the magnetic susceptibility of the material can be derived through graphing  $F$  vs.  $(H_0^2 - H_t^2)$ , in which  $F$  represents force,  $H_0$  is the magnetic field at the bottom of the sample, and  $H_t$  is the magnetic field at the top of the sample.

$$F = \frac{m_s}{2L} \chi (H_0^2 - H_t^2)$$

The slope of such a graph will be equal to the magnetic susceptibility,  $\chi$ . For ease in calculation and measurement, a proportional value of magnetic susceptibility was calculated through graphing differences in the initial mass and the mass in the applied external magnetic field vs. the current squared. The differences in masses were used in order to better account for the drift of the measurements. To observe temperature dependence, the magnetic susceptibilities were calculated at two different temperatures: 298 K at room temperature and 77 K in liquid nitrogen.



**Figure 10: Magnetic susceptibility measurements taken at room temperature (a, b, c) and liquid nitrogen temperature (d, e, f) for anhydrous, monohydrate, and pentahydrate crystals, respectively**



**Figure 11: Curie-Weiss law relationship lines of a) anhydrous, b) monohydrate, and c) pentahydrate, respectively**

The graphical interpretation of the Curie-Weiss law shows the Curie-Weiss temperatures. Since all y intercepts of these graphs in Figure 11 were negative, the anhydrous, monohydrate, and pentahydrate copper(II) sulfates are all considered antiferromagnetic. However, a clear trend is displayed through the Curie-Weiss temperatures that shows that as more water molecules are added to the copper(II) sulfate, the materials become less antiferromagnetic. In fact, this decrease in antiferromagnetic behavior is so dramatic that copper(II) sulfate pentahydrate is close to being paramagnetic: a completely different magnetic behavior than anhydrous copper(II) sulfate. The results of the magnetic susceptibility experiment show that the addition of water molecules to the structure affects magnetic properties.

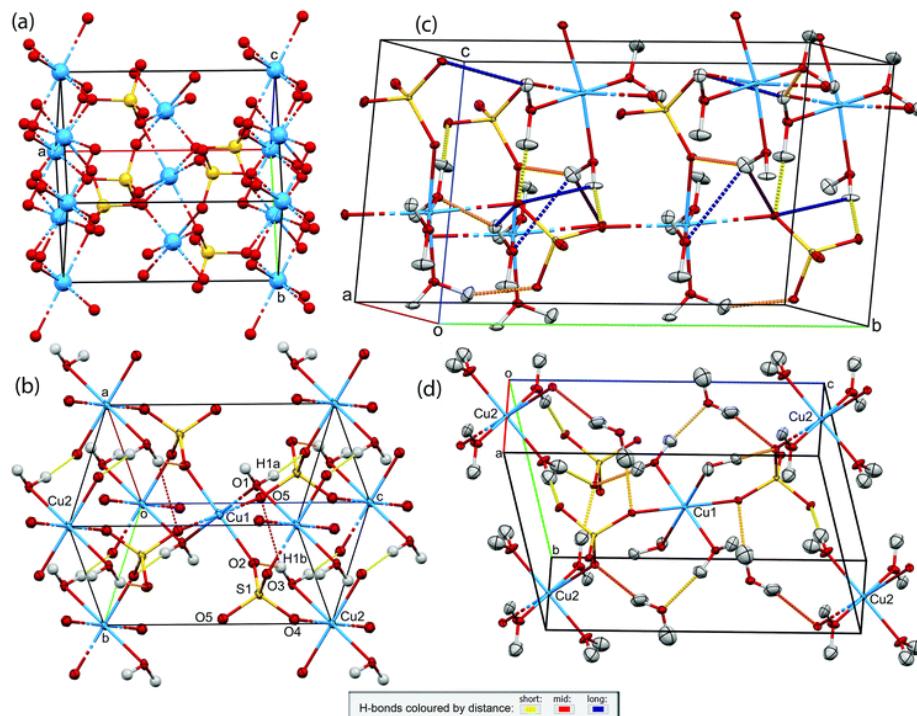
## V. Conclusion

### A. What Do the Results Indicate?

The results indicate that changing the crystal structure by adding waters of hydration affects electronic and magnetic properties. Since water has all paired electrons, it doesn't affect ESR. Water is also known to be diamagnetic, meaning it always has a zero or slightly negative magnetic susceptibility. Any change in magnetic properties comes almost entirely from changing the position of the copper(II) and sulfate ions within the crystal. Using XRD, it was also shown that the crystals were distinct, and only one derivative of the crystal was present in each sample. What this indicates is the position of the ions within a crystal has an important effect on magnetic properties, because changing the crystal structure changes the ESR data and magnetic susceptibility. Additionally, it seems like the addition of water molecules made the crystal more paramagnetic rather than antiferromagnetic.

### B. Areas of Future Research

An obvious area of future research would be to test other hydrates of copper sulfate and see if the same pattern is present among all of the hydrated crystals of Copper(II) Sulfate. Another area of future research would be to test if adding waters of hydration affects the crystal structure of other inorganic crystals. However, the area of future research we are most interested in is determining if there is a way to predict magnetic properties from the structure of a crystal. Our experiment indicated that the structure has an effect on magnetic properties, we now want to know if that can be quantified and used to predict the magnetic properties of other crystals.



**Figure 12: Structures of copper(II) sulfate derivatives: (a) anhydrous, (b) monohydrate, (c) trihydrate, and (d) pentahydrate<sup>8</sup>**

In the future, we would perform x-ray crystallography on our samples to better understand how the ions within them are spatially arranged. X-ray crystallography is similar to the powder x-ray diffraction we did in our study, however, instead of a powder, a single crystal is used. This allows for the computation of the unit cell. From this data, we can see if there is a correlation between crystal structure and magnetic properties. We would have to test a much wider range of crystals than in this experiment to see if there is a correlation.

## Acknowledgements

We would like to thank our advisor Dr. Barry Luokkala, teaching assistant Annie Cheng, and faculty advisor Michael Saavedra for guiding us throughout all of the research, helping us learn how to use the equipment, and navigating the big ideas of the research. We would also like to thank the PGSS Campaign, Inc and the other sponsors of PGSS for making this opportunity possible.

## References

<sup>1</sup> J. S. Wells and L. M. Matarese, "Electron Spin Resonance in Single Crystals of Anhydrous Copper Sulfate", *The Journal of Chemical Physics*, <https://tf.nist.gov/general/pdf/412.pdf>

<sup>2</sup> Hippocampus OU, <https://hippocampus.ee/product/copper-sulphate-pentahydrate-feed-grade/>

<sup>3</sup> David M. Hanson, Erica Harvey, Robert Sweeney, Therea Julia Zielinski, "An Electron has an Intrinsic Spin Angular Momentum". *Chemistry Libre Texts*, 8.4

[https://chem.libretexts.org/Bookshelves/Physical\\_and\\_Theoretical\\_Chemistry\\_Textbook\\_Maps/Physical\\_Chemistry\\_%28LibreTexts%29/08%3A\\_Multielectron\\_Atoms/8.04%3A\\_An\\_Electron\\_Has\\_an\\_Intrinsic\\_Spin\\_Angular\\_Momentum](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Physical_Chemistry_%28LibreTexts%29/08%3A_Multielectron_Atoms/8.04%3A_An_Electron_Has_an_Intrinsic_Spin_Angular_Momentum)

<sup>4</sup> The Editors of Encyclopaedia Britannica, "X-ray diffraction". *Encyclopedia Britannica*, 31 Mar. 2023, <https://www.britannica.com/science/X-ray-diffraction>. Accessed 4 August 2023.

<sup>5</sup> "Ferro-, Ferri, and Antiferromagnetism", *Chemistry Libre Texts*, 6.8  
[https://chem.libretexts.org/Bookshelves/Inorganic\\_Chemistry/Book%3A\\_Introduction\\_to\\_Inorganic\\_Chemistry\\_%28Wikibook%29/06%3A\\_Metals\\_and\\_Alloys-\\_Structure\\_Bonding\\_Electronic\\_and\\_Magnetic\\_Properties/6.08%3A\\_Ferro-\\_Ferri-\\_and\\_Antiferromagnetism](https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Book%3A_Introduction_to_Inorganic_Chemistry_%28Wikibook%29/06%3A_Metals_and_Alloys-_Structure_Bonding_Electronic_and_Magnetic_Properties/6.08%3A_Ferro-_Ferri-_and_Antiferromagnetism)

<sup>6</sup> Eve M. Mozur and Ram Seshadri, "Methods and Protocols: Practical Magnetic Measurement" *Chemistry of Materials* 2023 35 (9), 3450-3463  
DOI: 10.1021/acs.chemmater.3c00297

<sup>7</sup> "Electron Spin Resonance", *Chemistry Libre Texts*, 7.6  
[https://chem.libretexts.org/Courses/University\\_of\\_California\\_Davis/UCD\\_Chem\\_107B%3A\\_Physical\\_Chemistry\\_for\\_Life\\_Scientists/Chapters/7%3A\\_Spectroscopy/7.6%3A\\_Electron\\_Spin\\_Resonance](https://chem.libretexts.org/Courses/University_of_California_Davis/UCD_Chem_107B%3A_Physical_Chemistry_for_Life_Scientists/Chapters/7%3A_Spectroscopy/7.6%3A_Electron_Spin_Resonance)

<sup>8</sup> Mukaila A. Ibrahim, René T. Boéré The copper sulfate hydration cycle. Crystal structures of CuSO<sub>4</sub> (Chalcocyanite), CuSO<sub>4</sub>·H<sub>2</sub>O (Poitevinite), CuSO<sub>4</sub>·3H<sub>2</sub>O (Bonattite) and CuSO<sub>4</sub>·5H<sub>2</sub>O (Chalcanthite) at low temperature using non-spherical atomic scattering factors, *Royal Society of Chemistry*

## **APPENDIX**



Angela Abraham State College Area High School IU 10	Dinuk DeAlmeida North Allegheny SHS IU 3	Yeana Kim Warwick High School IU 13
Vanesa Aguay Wyomissing Area High School IU 14	Renuk DeAlmeida North Allegheny SHS IU 3	Trinity Kong Wyoming Seminary IU 18
Kofi Anokye Wilson High School IU 14	Suraj Dumasia Manheim Township High School IU 13	Selina Lin Hershey High School IU 15
Sanjana Balaji Garnet Valley High School IU 25	Isabelle Ehrensberger Ridgway Area High School IU 9	Dorothy Liu Exeter Township Senior High IU 14
Grace Barlett Allegheny-Clarion Valley HS IU 6	Harry Feng Wyoming Seminary IU 16	Yanwei Liu Hershey High School IU 15
Jonathan Barsotti Friends Central School IU 25	Cassidy Griffith Peters Twp High School IU 1	Helen Mao North Allegheny SHS IU 3
Noah Beckish North Pocono High School IU 19	Megan Grolemund Freeport Area High School IU 28	Edward Mei North Allegheny SHS IU 3
John Beeson Pittsburgh Central Catholic HS IU 7	Izabella Gurreonero Norristown Area High School IU 23	Olina Mukherjee Allderdice High School IU 2
Vipin Bhat Allderdice High School IU 2	Thomas Hasty Springfield Township HS IU 23	Sanjay Nair Central Bucks South HS IU 22
Noah Breithaupt Moravian Academy IU 20	Jessica Hou J.R. Masterman High School IU 26	An Ben Nguyen Montoursville Area High School IU 17
Carter Brown Pottsville Area High School IU 29	Zheyu Ray Hu Conestoga High School IU 24	Kaila O'Connor Jim Thorpe Area High School IU 21
Damien Busche Peters Twp High School IU 1	Viki Huang New Brighton Area High School IU 27	Parth Parikh McDowell High School IU 5
Karina Chan-van der Helm Springside Chestnut Hill Academy IU 26	Namya Jindal Dallastown Area High School IU 12	Carlos Peterson George Washington Carver HS IU 26
Hannah Cudzil Oswayo Valley School IU 9	Gayath Karunaratne Westmont Hilltop High School IU 8	Lucas Philip Germantown Academy IU 26

Kamya Rajesh North Allegheny SHS IU 3	Sooraj Tharumia Parkland High School IU 21	Wilson Zhang Conestoga High School IU 24
Nidhi Ram Lower Moreland High School IU 23	Claire Thomas Lancaster Mennonite School IU 13	
Marbella Ramos Guardado Hazleton Area Academy of Sciences IU 18	Diya Thomas Central Bucks South HS IU 22	
Lihini Ranaweera Seneca Valley High School IU 4	Antony Tomy Parkland High School IU 21	
Krishna Ravi McDowell High School IU 5	Shivani Umesh Seneca Valley High School IU 4	
Brighton Risch Great Valley High School IU 24	Mackenzie Vasbinder Mifflinburg Area High School IU 16	
Carter Rogers Muncy Jr./Sr. High School IU 17	Harihar Viswanathan Shady Side Academy IU 7	
Ryan Sajwan Lehigh Valley Academy IU 20	Darren Wang Hampton High School IU 3	
Rachael Saldubehere Mifflin County High School IU 11	Ronald Wang Indiana Area Senior HS IU 28	
Keira Seidman Bensalem High School IU 22	Meghan Wong Scranton High School IU 19	
Prakhar Singh Clarion Area High School IU 6	Megan Xie Lower Moreland High School IU 23	
Savannah Stackhouse Central Cambria High School IU 8	Eric Xu Conestoga High School IU 24	
Arjun Suryawanshi Unionville High School IU 24	Valerie Xu Conestoga High School IU 24	
Alem Tesfaye Central York High School IU 12	Felix Yang Mt. Lebanon High School IU 3	

John Alison  
Department of Physics  
Carnegie Mellon University

Niharika Botcha  
Department of Chemistry  
Carnegie Mellon University

Benjamin Campbell  
Department of Engineering  
Robert Morris University

Chayanika Chaudhuri  
Physics Department  
Robert Morris University

Hael Collins  
Department of Physics  
Carnegie Mellon University

Carrie Doonan  
Department of Biological Sciences  
Carnegie Mellon University

Lynley Doonan  
Department of Biological Sciences  
Carnegie Mellon University

Emily Drill  
Department of Biological Sciences  
Carnegie Mellon University

Doug Goetz  
Department of Chemistry  
University of Colorado

David Handron  
Department of Mathematics  
Carnegie Mellon University

Kerry Handron  
Physics and Astronomy  
Public VR/ Allegheny Observatory

Barry Luokkala, PGSS Program Director  
Department of Physics  
Carnegie Mellon University

Brooke McCartney  
Department of Biological Sciences  
Carnegie Mellon University

Andrew McGuier  
Computer Science  
Echo Gate Tech LLC

Natalie McGuier  
PGSS Asst. Program Director  
Department of Biological Sciences  
Carnegie Mellon University

Prahlad Menon  
Swanson School of Engineering  
University of Pittsburgh

Beka Modrekiladze  
Department of Physics  
Carnegie Mellon University

Michael Saavedra  
Department of Physics  
Carnegie Mellon University

Neil Simonetti  
Computer Science and Mathematics  
Bryn Athyn College

Stephanie Tristram-Nagle  
Department of Physics  
Carnegie Mellon University

John Urbanic  
Pittsburgh Supercomputing Center  
Carnegie Mellon University

Russell Walker  
Department of Mathematics  
Carnegie Mellon University



**Director of Resident Life**

Cameron Blackwood-Short  
Carlow University

**Teaching Assistants/Counselors/Lab Assistants**

Matthew Abode  
Biology  
Bucknell University

Annie Cheng  
Physics  
University of Pennsylvania

Ashley Driesbach  
Chemistry  
Penn State University

Christian Guinto-Brody  
Physics  
University of Virginia

Nikki Holmes  
Biology  
Villanova University

Tripp Hoover  
Mathematics  
Carnegie Mellon University

Liam Hower  
Computer Science  
Carnegie Mellon University

Lora Kallenberg  
Chemistry  
Carnegie Mellon University

Krystof Purtell  
Computer Science  
University of Pennsylvania

Erin Smith  
Chemistry Storeroom Attendant  
Carnegie Mellon University

Michael Sobol  
Biology  
University of Pittsburgh

Léa Tonnerre  
Mathematics  
University of Montana

Julia Wang  
Chemistry  
Carnegie Mellon University

Haley Williams  
Physics  
Carnegie Mellon University