MURJ

**Massachusetts Institute of Technology
Undergraduate Research Journal**

Features p.11

Quantum Computing @ MIT

News p.9

Mucus's Ability to Suppress

Microbes

BAYER | Life Hub Boston

#AdvancingInnovation

Life Hub Berlin · Life Hub Lyon · Life Hub California · Life Hub Boston · Life Hub Osaka · Life Hub Singapore

LifeHub Boston
Research

An innovation space by Bayer that brings the Boston scientific community together to network, learn, and collaborate.

Visit website ⬈

You're not following us yet?

+ Follow ···

lifehubboston.bayer.com ;)

Scan me!

| All | Images | Documents | Videos |

Sort by: Top ▾

Home
About
Ads

Check out our current events!

LifeHub Boston
426 followers
21h

Our community gatherings are open to everyone! Interested in Agriculture? Check out AgConnect! Maybe Pharma? Patient Voice is for you! Can't decide? We can't wait to see you at both!

Reduced Crop Loss
- Enabled by improved tolerance
- Reduces crop loss conditions
- Annual yield range from

Precision

👍 6

👍 Like   💬 Comment   ↪ Share

Be the first to comment on this

LifeHub Boston
426 followers
3h

Our Co-Creation Sprints are an important event for both Bayer professionals, and our LifeHub network! Being able to design solutions to big questions together is one of our favorite parts of working with the Boston ecosystem!

Check out our video

👍 Like   💬 Comment   ↪ Share

Be the first to comment on this

Highlights

Join our Mailing List!

Monthly newsletters
Event updates
Latest news from the Hub

Network with us!

We want to plug into the brightest minds in the Boston ecosystem, and we can't wait to meet you next. Will we see you at our next event?
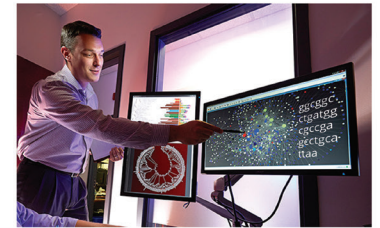
Check out our global network!

With locations in:
Berlin, U.K., Osaka, Lyon, California, and Singapore

innovate.bayer.com!

Contents

December 2019

Dear MIT community,

We are excited to present the 38th issue of the MIT Undergraduate Research Journal. Our emphasis on undergraduate research is a unique part of MIT culture. This year is the 50th anniversary of the Undergraduate Research Opportunities Program (UROP), which provides support and opportunities in research for 90% of MIT undergraduate students during their time here. We would like to acknowledge the thousands of passionate and talented students across campus working to further knowledge in their fields, as well as the mentors that dedicate their time to training the next generation of researchers. The articles published in this issue are representative of the high quality of undergraduate research that takes place every day.

This issue features reporting on topics as diverse as controversy over ethical mouse handling, to the correlation between college students' grades and adequate sleep, to a satellite designed by a team of MIT researchers that has uncovered an unexpected number of planets in our solar neighborhood. Also included is an exploration of the past, present, and future of quantum computing at MIT and a look at synthetic biology and the MIT iGEM team. Original research published in this issue spans the fields of political science, biological engineering, physics, and materials science.

Biannual publication of this journal is the product of hard work, collaboration and commitment by MURJ staff members. We would like to thank our editorial board and contributors for their time and hard work this semester. In addition, we would like to thank all the undergraduates who shared their research with

us and the greater MIT community. For previous issues of the MIT Undergraduate Research Journal, please visit our website at murj.mit.edu.

If you are interested in contributing to future issues of the MIT Undergraduate Research Journal, we would be delighted to have you. Please contact murj-officers@mit.edu if you have any questions or comments.

Best,

Allison Huske
Co-Editor-in-Chief

Natasha Joglekar
Co-Editor-in-Chief

Vaibhavi Shah
Co-Editor-in-Chief

SPACE

Exoplanet probe team awarded NASA's Silver Achievement Medal

MIT-led Transiting Exoplanet Survey Satellite (TESS) discovers more planets in our solar neighborhood than previously anticipated, receives NASA honor for "stellar achievement"

Launched in April 2018, the **Transiting Exoplanet Survey Satellite (TESS),** set out to search for worlds beyond our own—and so it did. In just over a year of operation, TESS has already found 29 planets and more than a thousand candidates, far exceeding its initial goals. Now, its search continues.

In recognition of their achievements, the TESS team was awarded NASA's Silver Achievement Medal, the agency's second-highest honor for government and non-government individuals or teams. The medal was presented on September 5 during a ceremony at NASA's Goddard Space Flight Center.

Every year, NASA confers the prestigious award for "a stellar achievement that supports one or more of NASA's core values, when it is deemed to be extraordinarily important and appropriate to recognize such achievement in a timely and personalized manner." Around 250 MIT scientists and engineers were recognized as medal recipients.

"The NASA Silver Achievement Medal recognizes the revolutionary impact that TESS is now having on the emerging field of exoplanets, as well as TESS's revealing of exciting new insights in stellar and extragalactic astrophysics," the mission's principal investigator George Ricker said. "The members of the TESS science and engineering teams can rightly be proud of the marvelous instrument which they have brought into operation."

TESS is an MIT-led NASA mission whose primary goal is to detect planets within our solar neighborhood. Before TESS, the Kepler/K2 satellite was the most productive, discovering 2,737 exoplanets before it was decom-
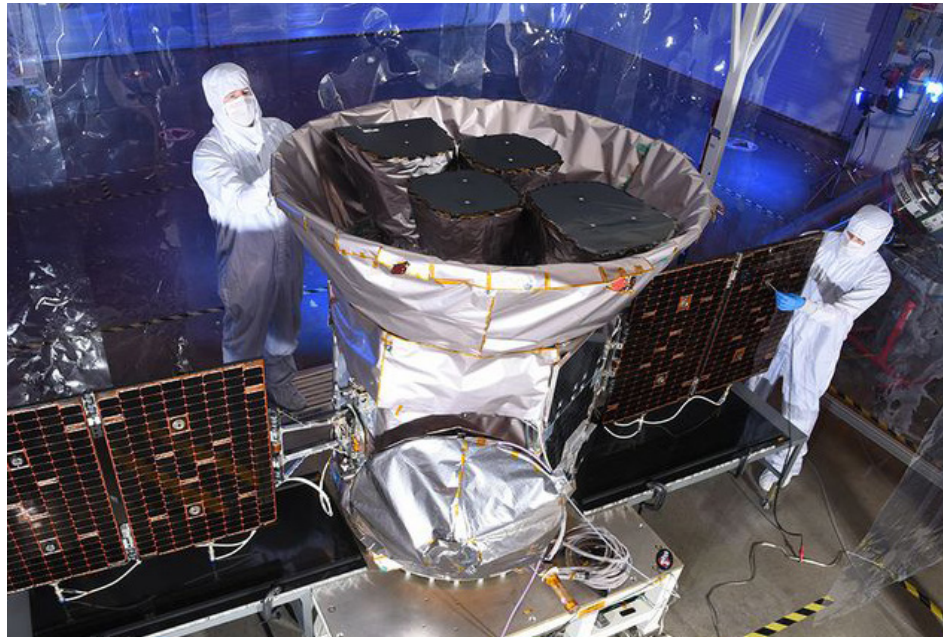
Technicians inspect the **TESS satellite** before launch in April 2018. (Photo: Northrop Grumman)

missioned in November 2018 after almost 9 years of operation. Building upon Kepler's pioneering success, TESS is expected to find more than 20,000 exoplanets, of which upwards of 50 are Earth-sized.

To detect extrasolar planets, both TESS and Kepler/K2 use transit photometry—the most effective and sensitive detection technique to date. The satellites monitor extremely slight dips in a star's brightness when a planet "transits" or passes in front of a star for a few hours. From this, data about the planet's size and orbital parameters are obtained.

TESS has some key features distinct from Kepler, however. Unlike Kepler which focused on a single patch of sky around the constellation Cygnus, TESS's search covers almost the entire sky—an area around 350 times larger.

In addition, whereas Kepler's plan-

ets are 300 to 3,000 light-years away, TESS's will be ten times closer. This way, its discoveries will be amenable to follow-up observations by other ground- and space-based facilities.

In its search, TESS uses four identical cameras developed by the MIT Kavli Institute and the MIT Lincoln Lab. But because it is only a survey satellite, it has to rely on external instruments, such as the forthcoming James Webb telescope, for further characterization of planet masses, densities, and atmospheric compositions.

Although initially slated for a two-year prime mission, the TESS mission was extended until 2022. Until then, it shall persist on its quest—sifting through the light of distant stars, fueled by humanity's collective longing to find a world like ours.

—Hillary Diane Andales

Seeking innovators

Solving the toughest health challenges is what drives us. AbbVie is a global biopharmaceutical company passionately committed to the possibilities of a healthier world. We're innovators who won't stop until we find the answers to make a remarkable impact on people's lives. Together, we work in a collaborative environment where scientists and commercial teams rally behind a common mission.

Learn more at
abbvie.com/careers

abbvie

People. Passion. Possibilities.®

LIFESTYLE

Want Better Grades? Get Some Sleep!

New research on test scores and sleep has important implications for college students

There's a reason why nobody tells college students they need to get less sleep. A recent MIT-Harvard study by Kana Okano, Jakub R. Kaczmarzyk, Neha Dave, John D. E. Gabrieli, and Jeffrey C. Grossman found that increased duration, quality, and consistency of sleep are associated with better grades.

Although sleep-related research has been done in the past, these findings suggest novel ways for students to improve their academic performance. In contrast to prior studies, which have found that increased sleep quality and duration the night before an exam was associated with better performance, the data collected by Okano et al. showed a correlation between test results and sleep for an entire month before a given midterm. The researchers wrote, "Rather than the night before a quiz or exam, it may be more important to sleep well for the duration of the time when the topics tested were taught."

Another surprising finding may help explain gender disparities in academic performance. Previous studies have found that females consistently out-perform males in higher education; however, after sleep patterns were statistically equated, Okana et al. discovered that there was no longer a significant female advantage in exam scores. Regarding this, the researchers remarked, "[I]t may be

> *"Rather than the night before a quiz or exam, it may be more important to sleep well for the duration of the time when the topics tested were taught."*

especially important to encourage better sleep habits in male students (although such habits may be helpful for all students)."

Despite these intriguing data, the researchers acknowledge that their results may not be fully conclusive. The study was conducted on 88 MIT students enrolled in 3.091 (Introduction to Solid-State Chemistry), 85 of whom were freshmen — in short, not reflective of the general population, nor even the general college population. In addition, the students' sleep was assessed using FitBits, but there is no indication whether the "sleep quality" metric provided by the activity trackers reflected actual sleep quality. Furthermore, the data reveals only correlation, not causation: though early bedtimes were associated with early wake-up times, which were in turn associated with better academic performance, the researchers were reasonably hesitant to declare that increase sleep was responsible for this improvement since students with better study habits would likely have both better sleep patterns and test scores. Nevertheless, Okana et al. remain optimistic that future research will be able to resolve the questions raised by their study.

— Teresa Gao

PHYSICS

MIT Professor Emeritus receives $3 million prize for work on supergravity

Daniel Freedman wins Special Breakthrough Prize in Fundamental Physics for solving a longstanding problem in physics

In 1976, three theorists realized Einstein's dream: to unify all four fundamental forces (strong force, weak force, electromagnetism, and gravity) under a single mathe-matical framework. They birthed a revolutionary idea, dubbed "supergravity," pivotal in the development of physical theories after it.

Just this year, more than four decades since supergravity's inception, its originators were

Daniel Freedman (right) along with collaborators **Peter van Nieuwenhuizen** (left) and **Sergio Ferrara** (center) are this year's recipients of the Special Breakthrough Prize for their work on supergravity. (Photo: CERN)

announced as winners of the Special Breakthrough Prize in Fundamental Physics. Among them is Daniel Freedman, MIT Professor Emeritus in both the Mathematics and Physics Departments.

> *"Supergravity paved the way for string theory, currently the leading candidate for the theory of everything."*

Freedman will share the hefty prize of $3,000,000, considered the most generous in the world, with collaborators Peter van Nieuwenhuizen of Stony Brook University and Sergio Ferrara of CERN. "This one takes the cake—it is the cap of my long career," he says to MIT News.

Supergravity attempts to solve one of physics' most enduring conundrums: creating a coherent mathematical framework from which all four fundamental forces emerge. The Standard Model—pillar of particle physics—had so far been successful at unifying all forces but one. According to the model, interactions via the fundamental forces are mediated by particles, like how the photon mediates electromagnetism. Yet, gravity remained absent from its picture. That was, until supergravity.

To integrate all the forces, supergravity incorporates the concept of supersymmetry into gravity, hence the name. Supersymmetry posits that all known particles in the Standard Model must have their own "superpartners" by virtue of some symmetry in their spins. Now, the trio's idea adds gravity into the picture, predicting a mediator particle called the graviton and its superpartner called the gravitino.

A mathematical tour-de-force, supergravity paved the way for string theory, currently the leading candidate for the theory of everything. String theorist and member of the Breakthrough Prize selection committee Andrew Strominger affirms in an interview with Nature that their award is well-deserved, describing their work as "being transcendently important in the development of physics for the past 40 years."

Other physicists declare the contrary, however, citing the theory's lack of empirical support despite decades of attempts to find even the slightest hint of a supersymmetric particle. Quantum gravity physicist Sabine Hossenfelder says to Nature that "the award should be for pure mathematics because this is not physics."

Although supergravity is yet to find its prized particles, physicists agree that it is equal parts imaginative and influential. Yuri Milner, one of the founders of the Breakthrough Prize, said in a statement: "Supergravity has inspired physicists for decades and may contain deep truths about the nature of reality." Until experiments definitively rule it out, supergravity shall remain a compelling description of reality.

— Hillary Diane Andales

BIOLOGY

Mucus's Ability to Suppress Microbes

The Ribbeck Lab at MIT is infamous for being a mucus research hub. Yes, mucus, the biomaterial that we are all familiar with during the cold season. Yet, many don't know that around a liter of mucus is produced by your body every day in order to coat the multitudes of tracts throughout your body systems. Many think about mucus with disgust and couldn't begin to imagine what it would be like to study it year-round, not to mention recognize how essential it really is.

Katharine Ribbeck's lab just published a paper in Nature Microbiology that explains how the chemical make-up of mucus actually protects from harmful bacteria. In particular, there are compounds in mucus calls mucins that are long polymers with concentrations of sugars. Before this research, mucins had not been well characterized. Scientists previously hypothesized that mucus functions as a barrier to microbes, but the Ribbeck lab has now shown that that model is not entirely accurate. They instead observed that bacteria in mucus don't get stuck and immediately subdued.

Ribbeck and her lab tested the role of glycans in mucus by isolating the sugar compounds and exposing them to a pathogen common in cystic fibrosis patients and those with compromised immune systems, *Pseudomonas aeruginosa*. They observed the effect on the

> *"The findings suggest that mucus contains factors that regulate microbial behavior at the level of gene expression."*

formation of biofilms and discovered that the biofilms exposed to the mucus dissociated from their surface and 70% of the cells shifted to a planktonic phase. "Planktonic" describes a state where the bacteria are free moving. This result is in contrast to the previous model that suggested that clumps of microbes would form. The theory is that having microbes moving freely in the mucus is a much easier target for the immune system.

Then, to test if the mucus affects other virulence traits than just biofilms, the scientists measure the gene expression involved in infection. The results revealed that the mucus actually suppresses key phenotypic characteristics of bacteria, including quorum sensing. These two findings suggest that mucus contains factors that regulate microbial behavior at the level of gene expression.

The new study is the first ever to identify the sugar component of mucins and their antimicrobial behaviors. This new research reveals how these glycans prevent bacteria from forming biofilms by actively altering the way they communicate. The Ribbeck Lab's research shows again how mucus is not just a barrier to harmful particles, but is an important component of the immune system.

— Catherine Griffin

MISCELLANEOUS

Mouse-Handling: A Touchy Subject

April 24th marks World Day for Laboratory Animals, and with an estimated 95 percent of these animals used in laboratory research being rats or mice (NABR 2018), it's time we show a little extra appreciation. Mice have been vital in developing treatments for human diseases ranging from cystic fibrosis to Alzheimer's, and they also represent a pivotal model for new discovery in emergent research areas such as the brain and the microbiome. Nature news writer Tom Clarke has gone so far as to conclude, "A dog may be man's best friend but the humble mouse, Mus musculus, is certainly our greatest ally" (2002).

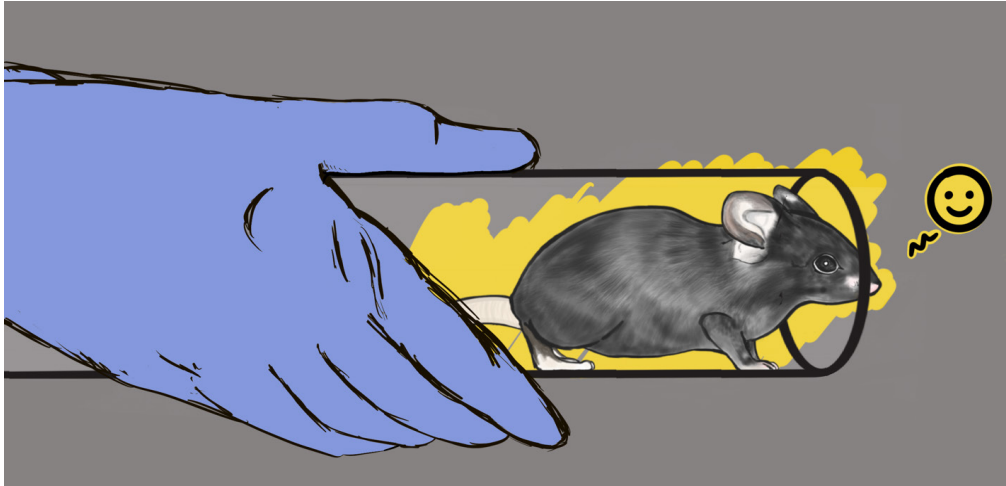You might have several ideas for mouse TLC. Perhaps a dollop

Image by Rachel Rock

of artisanal almond butter, a rub on the tummy, or a little toy to chew or toss about with cage mates? Or maybe even tickling? (That does work rather nicely for rats, after all). Yet, in all your tender thought, you've likely skipped over what your mice would appreciate most greatly of all—for you to stop dangling them by the tail!

Indeed, while numerous facilities across the US have taught "tailing" as the gold standard in mouse handling, a growing body of research suggests that this method is not only less-than-ideal, but also a significant stressor for the animals.

The research on tail handling and stress began with the Nature Protocol Paper published in 2010 by Hurst and West. In the paper, UK Principal Investigator Jane Hurst assessed the responses of male and female mice from three common laboratory mouse strains (BALB/c, C57BL/6, and CD-1) to three different handling methods: 'tail' (picked up and restrained by the base of the tail), 'cup' (scooped into gloved hand, then allowed to walk freely over the hands without direct physical restraint), or 'tunnel' (coaxed into a clear, acrylic home cage tunnel and

then, once inside, lifted without direct contact). Tail mice exhibited dramatically reduced voluntary interaction with handlers relative to cup and tunnel mice. Moreover, tail mice showed significantly elevated stress evidenced by increased defecation and urination during handling as well as anxiety-indicative behavior in a behavioral assay known as the Elevated Plus Maze.

Since then, Hurst's group has not only replicated this finding

"Handling-induced stress can weaken the reliability of behavioral studies."

(Gouveia and Hurst 2013), but expanded upon it powerfully. In 2017, Gouveia and Hurst showed that mice handled via tunnel as opposed to tail exhibited substantially improved performance on a simple task where mice discriminated between two different urine stimuli in successive trials. Tail mice exhibited dramatically reduced exploratory behavior and willingness to investigate test stimuli, and the mice only slightly improved their resultant poor performance with prior

familiarization to the test arena. This result is striking: such a finding suggests that handling-induced stress can weaken the reliability of behavioral studies.

Numerous other UK researchers have validated the use of non-aversive handling methods (cup or tunnel) as opposed to the tail handling technique. Ghosal S et al. (2015) found that mice cupped and fed a high-fat diet for three months exhibited improved glucose tolerance compared to tail-handled controls. Ono M et al. (2016) found tail-handling could elevate corticosterone (the mouse version of cortisol) in C57BL/6 mice. A study by Clarkson et al. (2018) showed that tunnel as opposed to tail handling improved sensitivity to a sucrose reward. Perhaps the most powerful proof of all, however, lies in practice: most UK facilities have replaced tail-handling with non-aversive practice.

If you'd like to give your mice a little "pick me up" and help them best help your lab in conducting valuable research, you can learn more here: https://www.nc3rs.org.uk/how-to-pick-up-a-mouse

— Rachel Rock

MURJ
Features

Quantum Computing @ MIT

By Francisca Vasconcelos

INTRODUCTION

Every school day, hundreds of MIT students, faculty, and staff file into 10-250 for classes, seminars, and colloquia. However, probably only a handful know that directly across from the lecture hall, in 13-2119, four cryogenic dilution fridges, supported by an industrial frame, endlessly pump a mixture of helium-3 and helium-4 gases to maintain temperatures on the order of 10mK. This near-zero temperature is necessary to effectively operate non-linear, anharmonic, superconducting circuits, otherwise known as qubits. As of now, this is one of the most widely adopted commercial approaches for constructing quantum processors, being used by the likes of Google, IBM, and Intel. At MIT, researchers are working not just on superconducting qubits, but on a variety of aspects of quantum computing, both theoretical and experimental.

In this article we hope to provide an overview of the history, theoretical basis, and different implementations of quantum computers. In Fall 2018, we had the opportunity to interview four MIT faculty at the forefront of this field — Isaac Chuang, Dirk Englund, Aram Harrow, and William Oliver — who gave personal perspectives on the development of the field, as well as insight to its near-term trajectory. There has been a lot of recent media hype surrounding quantum computation, so in this article we present an academic view of the matter, specifically highlighting progress being made at MIT.

QUANTUM COMPUTERS: A BRIEF HISTORY

The notion of a quantum computer was first introduced by Caltech Professor (and MIT alumnus) Richard Feynman in his 1959 "There's Plenty of Room at the Bottom" lecture, suggesting the use of quantum effects
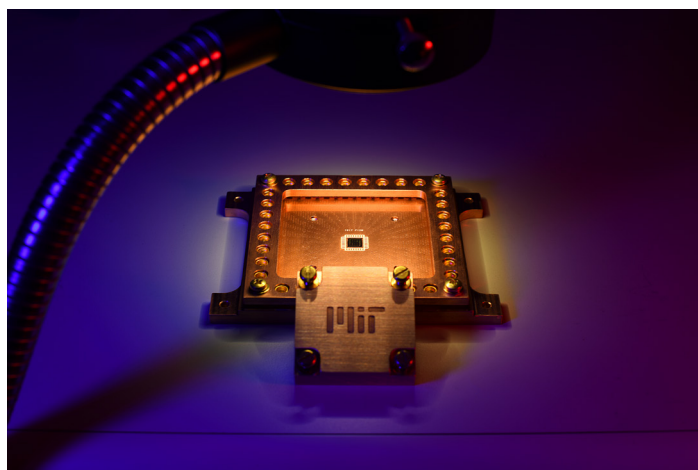
16-qubit superconducting quantum processor and its packaging, designed by researchers of the MIT Engineering Quantum Systems Group and fabricated at Lincoln Labs. [Credit: MIT News]
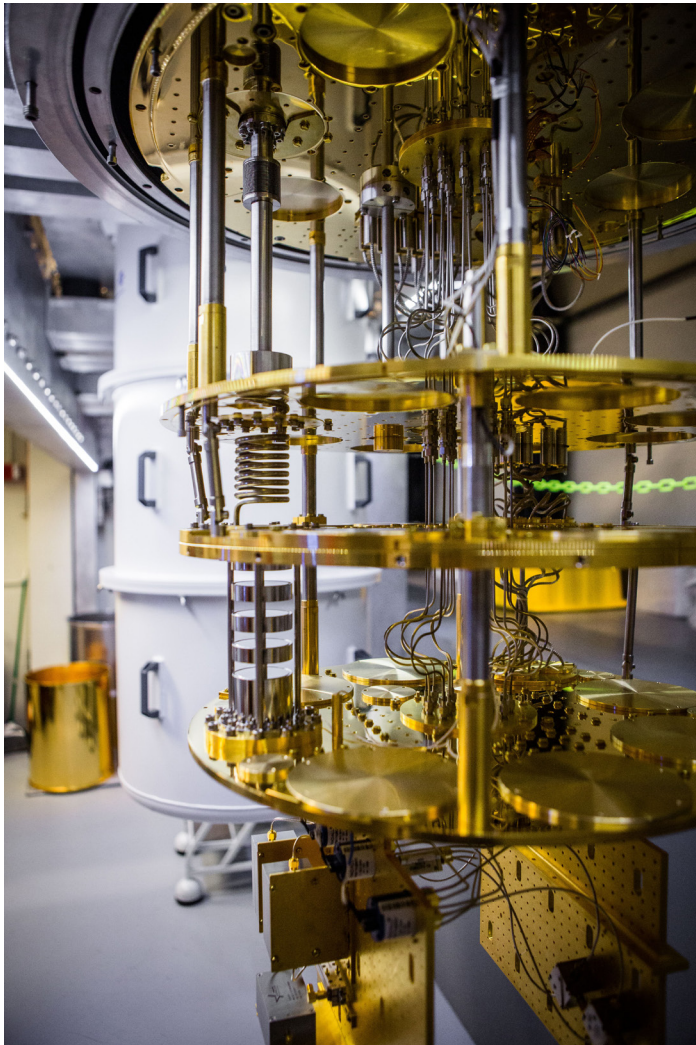
for computation. However, it was not until the late 1970s that researchers truly began exploring the idea. By May 1980, Paul Benioff — then a researcher at the Centre de Physique Théorique, CNRS — published the first article on quantum computation, "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines", in the Journal of Statistical Physics. In the same year, Russian mathematician and Steklov Mathematical Institute faculty, Yuri Manin published the book "Computable and Uncomputable", motivating the development of quantum computers. In 1981, MIT held the very first conference on the Physics of Computation at the Endicott House, where Benioff described computers which could operate under the laws of quantum mechanics and Feynman proposed one of the first models for a quantum computer. Clearly, MIT had roots in the field from its conception.

In 1992, David Deutsch and Richard Josza, proposed the Deutsch-Josza algorithm. Although the algorithm in

itself was not very useful, a toy problem some may say, it was one of the first demonstrations of an algorithm that could be solved far more efficiently on a quantum computer than on a classical (or non-quantum) computer. This idea, that quantum computers may exhibit a better computational complexity than classical computers, is known as quantum advantage. In the years that followed, more work was done in the domain of quantum algorithms and in 1994 one of the

Bluefors dilution refrigerator used in the Engineering Quantum Systems Group, of Prof. William Oliver, to keep superconducting quantum processors at near absolute zero temperatures. [Credit: Nathan Fiske]

biggest results in the field was achieved. Current MIT Professor of Applied Mathematics, Peter Shor, who was then working at Bell Labs, proposed the now well-known Shor's algorithm. This quantum algorithm reduced the runtime of integer number factorization from exponential to polynomial time. While this may

seem like a fairly abstract problem, the assumption that integer factorization is hard is the foundation of RSA security, the primary cryptosystem in use today. As the first algorithm to demonstrate significant quantum advantage for a very important problem without obvious connection to quantum mechanics (like quantum simulation), Shor's algorithm caught the attention of the government, corporations, and academic scientists. This created major interest in the field of quantum information and gave experimental researchers a strong justification to develop quantum computers from the hardware end.

The years following the announcement of Shor's algorithm witnessed significant developments in quantum computation, from both the theoretical and experimental perspectives. In 1996, Lov Grover of Bell Labs developed the quantum Grover's algorithm, which provides a quadratic speedup for search in unstructured problems, using a technique now known as amplitude amplification. That same year, the US Government issued its first public call for research proposals in the domain of quantum information, signifying the growing interest in quantum computation. Additionally, David P. DiVincenzo, then at IBM Research, listed the five main requirements to realize a physical quantum computer. Among these included the challenges of isolating quantum systems from their noisy environments and accurately controlling unitary transformations of the system. 1997 witnessed the first publications of papers realizing physical gates for quantum computation, based on nuclear magnetic resonance (NMR). Three out of the five authors on these transformative papers (Isaac Chuang, Neil Gershenfeld, and David Cory) are current or previous MIT faculty. Additionally, that year, two new physical approaches to quantum computing were proposed, making use of majorana anyons and quantum dots. In 1998, the University of Oxford, followed shortly after by a collaboration between IBM, UC Berkeley, and the MIT Media Lab, ran the Deutsch-Jozsa algorithm on 2-qubit NMR devices.[1] These served as the very first demonstrations of an algorithm implemented on a physical quantum computer. This was soon followed by the development of a 3-qubit NMR quantum computer, a physical implementation of Grover's algorithm, and advances in quantum annealing. Thus, by the end of the twentieth century, researchers demonstrated the physical potential of

quantum computing devices.

The turn of the century would mark a transition of focus towards improving and scaling these devices. The early 2000s alone witnessed the scaling to 7-qubit NMR devices, an implementation of Shor's algorithm which could factor the number 15, the emergence of linear optical quantum computation, an implementation of the Deutsch-Jozsa algorithm on an ion-trap computer, and the first demonstration of the quantum XOR (referred to as the CNOT gate). The following years consisted of similarly remarkable developments in and scaling of quantum computation technology. In fact, this growth has been so impressive, that it is reminiscent of an exponential Moore's-Law-type growth in qubit performance. This progress has resulted in excitement for quantum computation far beyond the realm of academia. As mentioned earlier, several large tech companies now have well-established research divisions for quantum technologies and the number of quantum startups seem to double each year. Furthermore, there has been a large increase in interest from both the government, with the approval of a $1.2 billion Quantum Initiative Act in 2018 – one of the only bipartisan legislative acts passed in recent memory — and the general public, with Google's recent announcement of quantum supremacy. However, with this increased interest and the desire for easily-approachable explanations to complex research comes the tendency to hype the current state of the field. Thus, we interviewed four faculty at the forefront of academic research in quantum information and computation for their outlooks, to see if we could make sense of where all these quantum technologies are headed in the near- and long-term.

PROF. WILLIAM OLIVER – SUPERCONDUCTING QUANTUM PROCESSORS

William Oliver is the current Associate Director of the MIT Research Laboratory of Electronics (RLE), an Associate Professor of EECS, a Physics Professor of the Practice, and a Fellow of the MIT Lincoln Laboratory. He is also a Principal Investigator in the Engineering Quantum Systems Group (MIT campus) and the Quantum Information and Integrated Nanosystems Group (MIT Lincoln Laboratory), where he leads research on the materials growth, fabrication, design, and measurement of superconducting qubits.

How would Professor Oliver explain his research to a non-expert?

When we asked Professor Oliver to explain his research in general terms, he described superconducting qubits as "artificial atoms." Like natural atoms, they have discrete quantum energy levels, in which transitions can be driven. One key difference, however, is that these artificial atoms are macroscopic electrical circuits comprising an Avogrado's number of actual atoms. Given that a qubit is a circuit, its energy levels can be engineered to be more optimally suited for quantum computation.

Prof. William Oliver, PI of the MIT RLE Engineering Quantum Systems group. [Credit: Nathan Fiske]

At a high level, a superconducting qubit is an LC-circuit — an inductor and capacitor in parallel — like you might see in an E&M course, such as 8.02. This type of circuit is a simple harmonic oscillator with a harmonic potential, meaning an equal energy spacing between all the energy. However,

quantum bits are generally built from the ground and first excited energy levels, corresponding to the two possible bit states (0 and 1). In order to isolate these two lower energy levels from higher energy levels, quantum engineers make use of Josephson junctions as their inductors. At superconducting temperatures, Josephson junctions act as non-linear inductors, introducing an anharmonicity to the potential and creating unequal spacings between different energy levels.

Researchers in the Engineering Quantum Systems Group, of Prof. William Oliver, setting up racks to control and measure their superconducting qubits. [Credit: Nathan Fiske]

This allows the user to isolate the transition between the ground and first excited energy level from transitions to higher energy levels, solely by changing the microwave frequency at which the circuit is driven.

Where is the field of superconducting qubits going?

Superconducting qubits are currently in the process of transitioning from a scientific/laboratory curiosity to a technical reality. Professor Oliver believes that researchers need to bring a lot more to the field, in order to actually realize this technology. It is not sufficient to just demonstrate a one-qubit or two-qubit gate, but instead it necessary to build a reproducible system. In this effort, a new discipline is emerging, called quantum engineering, which bridges the gap between quantum science and conventional engineering. It covers a wide range of fields, including physics, mathematics, computer science, computer architecture, analog and digital design, control theory,

digital signal processing, materials, fabrication, and more. In the long term, Professor Oliver believes all of these distinct domains need to come together to make superconducting qubit technology a viable technology.

According to Oliver, one of the holy grails of modern quantum engineering research is to demonstrate a logical qubit, or redundant set of qubits that achieve higher performance in aggregate than as individuals. To achieve this, each of the individual physical qubits need to have a sufficiently high operational fidelity, or ability accurately perform gate operations. Currently, many superconducting qubit groups can achieve single-qubit fidelities on the order of 99.95%. However, coupled qubits have fidelities in the range of 95-99%%, with only a few groups (including Oliver's) able to achieve more than 99%. Professor Oliver believes that in order to demonstrate a logical qubit and run error-correcting codes with reasonable overhead, those fidelities should get better (consistently above 99.9%, and the higher the better). He hopes that the surface code, a quantum error-correction scheme, will be implementable and demonstrated on superconducting devices of 17-49 qubits within the next 5 years. Demonstrations of quantum error correction are a key milestone in the development of universal quantum computers, because it enables resilience through redundancy, enabling larger systems.

The field is currently in the self-proclaimed NISQ, or Noisy Intermediate Scalable Quantum, Era in which there is access to small, noisy quantum processors. During the next 5-10 years, in parallel with the push towards making logical qubits and demonstrating quantum error-correction, Oliver believes that the field will also need to develop algorithms that make use of currently available NISQ devices. It is crucial to develop a quantum algorithm that gives a quantum advantage and addresses a useful, meaningful problem.

What kinds of problems will we use quantum computers to solve?

Although the main national security driver for quantum computers is Shor's algorithm and cryptanalysis, Professor Oliver does not believe this will be the main commercial driver. In fact, he finds the commercial drivers (quantum chemistry, quantum materials, and optimization),

much more exciting. "These are just the thorny, sticky hard problems that we can't directly solve with classical computers because the number of degrees of freedom is too large and the number of requirements are too large. We make approximations and, of course, we try our best to simulate them with classical computers, but for some problems, we have to approximate to a degree that the answer we get out is just too fuzzy or is not meaningful."

"Almost every [quantum] algorithm that we know of today is connected to a person at MIT in some way."

Professor Oliver provided us a few examples of applications that he is especially interested in and believes can be addressed with quantum computers. The first is the prediction of new materials that have certain functionalities, such as topological materials. In particular, he is hopeful that we could find a way to build materials that superconduct at room temperature. In the domain of quantum chemistry, Oliver is particularly interested in theoretical work performed at Microsoft with respect to the nitrogen fixation process. Currently, ammonia is manufactured using the Haber-Bosch process that was developed over 100 years ago, during WWI. Nitrogen fixation is a necessary step to create the fertilizer required to feed an ever-growing human population. However, it also requires extremely high pressure and temperature, contributing to an estimated 1-2% of worldwide energy usage. We do not currently know how to improve the Haber Bosch process, because we do not understand how the main enzyme in the process catalyzes ammonia. Professor Oliver pointed out, however, that bacteria have an enzyme that can that serves as a catalyst for nitrogen fixation using the small amounts of energy provided by their metabolism. Thus, he believes there is a more efficient way we could go about this process and hopes that we can

find it, using quantum computers to understand the chemical catalysis process.

Quantum computing at MIT.

In wrapping up our interview with Professor Oliver, we asked if there was anything that should be emphasized in this article. He felt that one of the main takeaways should be that "despite the hype, MIT has really been there from the beginning and pioneered a lot of work in this area. Almost every [quantum] algorithm that we know of today is connected to a person at MIT in some way." He gave specific examples of Professors Peter Shor, Aram Harrow, Ed Farhi, and Seth Lloyd. While these faculty may not have developed the algorithms they are famous for at MIT, they have all been MIT faculty at some point or are currently faculty. In addition to being a theory powerhouse, Oliver believes that MIT continues to excel in experimental work. Although his group focuses on superconducting qubits, he highlighted research on trapped ions, quantum optics, quantum communications, and quantum sensing. Professor Oliver believes that "it is more than just quantum computing. It is quantum information science and quantum engineering, and that includes sensing, communication and networking, and computing."

PROF. DIRK ENGLUND – QUANTUM PHOTONICS

Dirk Englund is an Associate Professor of Electrical Engineering and Computer Science at MIT and the Principal Investigator of the Quantum Photonics Laboratory. His research interests include quantum optics, precision measurement, and nanophotonics. ay Earth doesn't really need us.

How would Professor Englund explain his research to a non-expert?

According to Professor Englund, information is always encoded in some physical system, whether that be a small magnet or a charge. As information storage is shrunk to smaller and smaller length scales, the regime of information manipulation and storage becomes very small-scale. At this

Prof. Dirk Englund, PI of the MIT RLE Quantum Photonics Laboratory Research Group. [Credit: MIT RLE]

small scale, nature behaves according to the laws of quantum mechanics, which is quite different from how we see nature in our everyday lives. This in turn puts a limit on how much we can shrink devices. After 60 years of semiconductor scaling, we are now finally hitting those length scales. At tiny length scales, it is no longer possible to confine electrons. Through a phenomenon known as quantum tunneling, electrons can hop from one device to another and tunnel through barriers they are not supposed to. According to Englund, "if you are a pessimist, you might say this is the end of the scaling laws. We can no longer improve the performance of computers and so on that have changed our lives. But if you're an optimist you can say, maybe it's an entirely new chapter, where you look at the glass half-full rather than half-empty. And that chapter is the quantum information era, where rather than bemoaning the strange properties of nature at the atomic scale, you make use of them."

Through his own research, Professor Englund hopes to explore new opportunities that emerge when encoding information in quantum physical systems, particularly those based on photonics. "This could lead to computers and other kinds of information processing devices, as well as sensing devices that can do things that we do not know would be possible in a world governed by classical mechanics." In the 1950s, the first transistors were made out of silicon. It took a couple decades until these transistors were refined to the single transistor level and then could be replicated and multiplexed. Englund believes that, today, we are in the quantum information era at a level in which we can reliably create single "quantum transistors." However, they are still bulky and they are hard to manage. Through quantum photonics, researchers are learning how to put many of these quantum "transistors" on devices and how to scale them up.

What is the trajectory of quantum computation over the next few decades?

In 2016, MIT Professors Dirk Englund and Seth Lloyd organized a government sponsored workshop titled the "Future Directions of Quantum Information Processing: A Workshop on the Emerging Science and Technology of Quantum Computation, Communication, and Measurement." Several of the top scientists from the US and abroad gathered in Arlington, Virginia, to identify challenges and opportunities in quantum information processing for the following 5, 10, and 20 years. Professor Englund summarized the key takeaways of the workshop for us.

By the end of the 5 year timeline, it was anticipated that we will be in the, previously defined, NISQ Era and that we will have quantum systems too large for any classical computer to predict the behavior of. Although he believes they will have a big impact on computer science related complexity arguments, Englund's primary question is whether these systems will be useful for something practical. He cited a few particularly promising examples of current applications research that he was excited about. In terms of simulating other

quantum systems, like molecules, there is potential for better first principle structure design in materials development. In terms of optimization problems, there is potential to optimize logistics chains and other related problems. And finally, in terms of specialized algorithms, there are versions of algorithms, like Shor's, that have been proposed which could potentially run on NISQ quantum devices. Alongside all this progress in applications, Englund looks forward to progress in benchmarking studies, "which pitch quantum computers that are going to be imperfect and have limited numbers of qubits and gates against classical computers. I think that we can do this at all is amazing, because classical computers have been developed for much longer and they have way way way more money being poured into them. [The fact] that there is benchmarking on certain problems is really interesting and exciting."

Looking out to the 10 or 15 year time frame, Englund hopes that the noisy aspect of these intermediate quantum computers will be fixable using error correction. With this error correction, he claims we will have more confidence that error-corrected general-purpose quantum computers will have useful applications. With regards to Shor's algorithm, Professor Englund believes that there is a reasonable chance that before 2030 there will be a quantum computer which can break RSA encryption at a level not doable with classical computers, "let's say about 1800-bit encryption." Englund says that even if there is only a 5% chance that in the next 10 years a quantum computer will break RSA, we should be very worried. "5% is a lot. Especially if you don't know if somebody has that computer."

Thoughts on media hype currently surrounding quantum computation? Will there be a quantum winter?

No matter what, Professor Englund believes that a lot of fantastic science will come out of all the

"There's a lot of interesting problems in-between, interesting questions that are going to be answered."

current research into quantum computing, both in terms of theory and experiment. With regards to commercial efforts in particular, he thinks it is fantastic that there is commercial activity. This allows systems to scale up much more quickly. That being said, Professor Englund believes it is still unknown what the most successful systems will be. "Some approaches will probably fold and some will continue to prosper. And then perhaps one can assign hype later, but it's tough to predict upfront what's going to be the best approach." Overall, however, he is optimistic that there will not be a general downturn on hype. "I think there will be some approaches that perhaps go away and some that will come up and some that will continue. That could make the field stronger as a result of it." He does not foresee a winter as dramatic as those experienced by the artificial intelligence community in the late twentieth century. In fact, Englund is confident there will continue to be big advances in experimental and theory efforts. "Some systems, once you investigate them for long enough, yeah maybe they won't scale. But that's ok. There are other systems that I think are going to go forward. I think there's several systems that I'm actually quite optimistic about. But it's not around the corner. I think some people in the media like to give the wrong perception that a general purpose quantum computer is a couple years away. It's probably still a decade away for a general purpose error-corrected computer. Maybe longer than that. But there's a lot of interesting problems in-between, interesting questions that are going to be answered. So, in my view, it is going to keep the field exciting, even before you have that ultimate goal of a general purpose error-corrected quantum computer. Which is fortunate for us, right, there are intermediate scale systems you can use before you have a full blown computer. That's super important."

PROF. ISAAC CHUANG – TRAPPED IONS

Isaac Chuang is a Professor of Physics and

Electrical Engineering, as well as the Senior Associate Dean of Digital Learning at MIT. He is the Principal Investigator of the MIT Quanta Research Group and a co-author of "Quantum Computation and Quantum Information," the primary textbook in the field.

How would Professor Chuang explain his research to a non-expert?

To a non-academic, Professor Chuang would describe his research as an attempt to build computers out of single atoms and single electrons. He says this is "cool" because these atoms obey special laws of physics, that can be used to solve problems that cannot otherwise be solved. However, if his middle-school daughter were to ask if she could use one, he would have to say "no sorry, not quite yet."

For someone with professional knowledge of science in general, Chuang says he is "trying to build a computer that can solve certain kinds of mathematical problems much faster than is possible with normal classical laws of physics, by using Schrodinger's equation." He claims that this is hard because quantum properties quickly disappear as we build larger systems. However, it is now believed that we can in fact keep those quantum properties intact, even while we scale our systems. In terms of exciting use cases of the technology, he cited the two major examples of Shor's and Grover's algorithms.

How did Professor Chuang get into quantum computing research?

According to Professor Chuang, there were two defining moments that lead him to quantum computing research. The first was rather early on, roughly 30 years ago as undergraduate at MIT, when he realized that he loved both computers and physics. "I wanted to do something that would let me play with both of these ideas. And it was a little sad for me to realize that I had to choose a major, which is either 6 or 8. I really wanted to do both." He started in 6 (EECS), but felt "a little bit bored and decided to learn about why things were the way they were." Eventually he

Prof. Isaac Chuang, PI of the MIT Quanta Research Group. [Credit: MIT Physics]

decided to double major in 8 (Physics).

Chuang claims that he only became interested in quantum computing per se when he read about Richard Feynman. He recounted digging through the MIT Physics Library to find Feynman's undergraduate thesis and reading any Feynman text he could lay his hands on. When Chuang started grad school, he decided that he did not want to do anything that was popular at the time. Instead, he "wanted to set out on [his] own and show that some of Richard Feynman's ideas would be feasible." He managed to convince a new professor, who had just joined the faculty, to let him "go off and wander around doing this, even though nobody was doing it." He believes there were only 6 people doing that kind of research in the world at the time. They "were all corresponding and saying this is a cool idea and thing to do, because there was no popularity in the concept of a quantum computer back then. There was Richard Feynman's 1985 article

in Optics News and all of us who knew about it would get excited about it, but that was it." Chuang wrote an article called "How to Build a Simple Quantum Computer" that was accepted to Phys Rev A and then Shor's algorithm hit the news. "One of my friends faxed me Peter Shor's article. And this is a preprint. No publication was able to publish as fast as the fax machines were going. I read it, I understood it, and I went around explaining to everyone I could possibly talk to. And I sat there going, I've gotta be able to realize this! After all, I had been one of the people working on the subject for fun. So then I set out to build it. And those were the early days. So, there was luck, but there was also this sense that I wanted to do something that wasn't possible. And I hope that's true of many of the undergrads today too."

"If we don't have insight into what other, different kinds of algorithms might exist... then the field won't go terribly far."

What were the expectations when Professor Chuang first joined the field of quantum computation?

Professor Chuang says that when he joined the field, it did not really exist. This meant that the few people working on the problem, at the time, set their own expectations and "they were a little crazy." In fact, he says that most people at the time expected quantum computing to be an utter failure. Even after Shor's algorithm came out and quite a few people were trying to rapidly realize the technology, they worried that quantum computing would follow a long history of alternate algorithm paths, like analog computation which failed largely because of noise. However, two advances came shortly thereafter, also by Peter Shor, which provided a great deal of hope for the field. The first was a seminal paper, which proved that there could be error correction for quantum systems.

"I wanted to do something that wasn't possible. And I hope that's true of many of the undergrads today too."

According to Chuang, "this is wild because a quantum system seems to be continuous valued and yet you can correct it." The second major breakthrough was a series of three seminal papers, which "showed why you can build an arbitrarily sized large-scale quantum computer out of competing parts that fail with certain probability, as long as that probability was lower than some threshold." This essentially reworked VonNeumann's theory of fault-tolerant computation for vacuum tubes into something useable for qubits. Professor Chuang claims that "if it were not for those papers [by Peter Shor], the whole field would have died." Furthermore, he believes Shor's ideas are why he would say that the expectations of the time were exceeded. "The biggest result for the first 20 years of quantum computing is largely that quantum computers are real. They are not just a theoretical abstraction. They can be realized in the laboratory."

What is the current state of quantum computing technology?

Professor Chuang believes we are solidly in the quantum "Vacuum Tube Era." He bases this claim off a great deal of reading he has done on the "Vacuum Tube Era" for classical computing. "If you go to the computer museum in Mountainview in California, you will see that companies actually built a large number of amazing vacuum tube computers, that did very sophisticated tasks. They were all made totally obsolete by the silicon transistor when it came along. But vacuum tubes actually went really far. And so, you might be tempted to think what we have today with quantum computers is already going far enough that you might call it a non-vacuum tube. However, they still fail, they fail with exceedingly

high probability, and we don't know how to step them up besides using error correction codes." Chuang notes, however, that there are theorists dreaming up alternate implementations of quantum computers which may overcome this challenge. In particular, the Microsoft Quantum research team has been looking into a method called topological quantum computers. Although extremely challenging to realize, such qubits would fail with much lower probability. "It is a very, very difficult route, but maybe something like that, someday, is something that I would call a quantum transistor."

What are the big challenges for quantum information?

Professor Chuang believes that the biggest problem currently facing the field is that there are really only two types of quantum algorithms. The first is the sub-exponential speed-up provided by versions of Grover's algorithm. The second is the exponential speedup provided by variants of Shor's algorithm, which all involve a Quantum Fourier Transform (QFT). There are several algorithms of this form, such as the Harrow-Hassidim-Lloyd algorithm, which gives exponential speedup for solving linear systems. However, Professor Chuang finds the fact that these algorithms all use the same structure as Shor's algorithm "rather frightening and disappointing." He believes that there should be other quantum algorithms, yet we have been struggling for over 15 years to try and discover them. "If we don't have insight into what other, different kinds of algorithms might exist and why they might be useful, then the field won't go terribly far."

Chuang also worries about the vast amount of speculation as to what might be feasible. He claims that people are coming up with all kinds of uses for quantum hardware, without knowing in principle how they ought to behave. To illustrate the point, he drew an analogy to machine learning in classical computing. "It just does well, people don't know why it does well. You can't prove any bounds on it. People are throwing the same logic at quantum computers and building quantum variational autoencoders

Prof. Aram Harrow of the MIT Center for Theoretical Physics [Credit: Justin Knight]

or other kinds of variational quantum algorithms, with no proofs. So, something has to progress along both of those lines."

PROF. ARAM HARROW – QUANTUM INFORMATION

Aram Harrow is an Associate Professor of Physics in the MIT Center for Theoretical Physics. His research interests include quantum algorithms, quantum information, quantum complexity theory, representation theory, and optimization.

How would Professor Harrow explain his research to a non-expert?

Generally speaking, Professor Harrow describes the goal of his research as "figure[ing] out what to do with a quantum computer and quantum communication network, if we had them." Additionally, he aims to use ideas from quantum information to improve our understanding
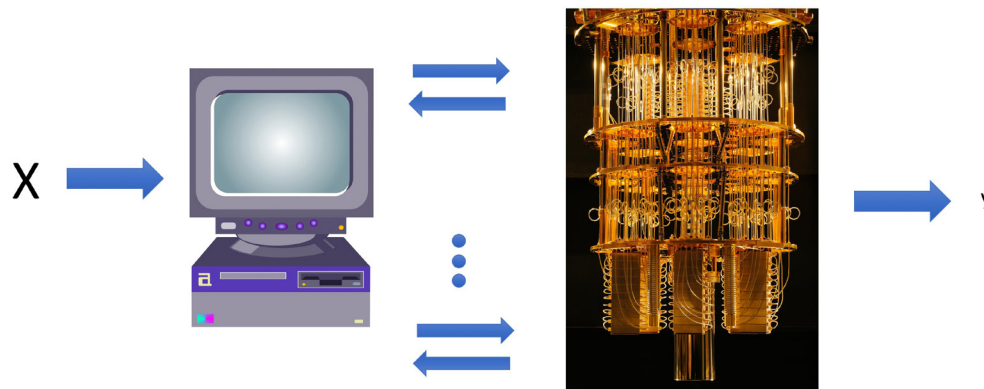
Image credit: IBM Corporation

A schematic of general hybrid classical-quantum algorithms, relevant to Prof. Aram Harrow's work on hybrid optimization algorithms. [Credit: Aram Harrow]

of physics and theoretical computer science. "Quantum mechanics was originally seen as a theory of physics, about small particles and how they behave. Recently, we have really begun to appreciate that it has a lot of implications for information and that this has emerged as a field in its own right. How should we update things like information, proofs, computing, and communication in light of quantum mechanics? And so, if you were to talk about what I did broadly, you could say that I just look at theoretical aspects of what quantum mechanics means for information."

What were the expectations when Professor Harrow first joined the field of quantum information?

Professor Harrow notes that the field was much smaller in the summer of 2000, when he started a UROP in MIT Professor Neil Gershenfeld's lab, working on an early NMR quantum computing experiment. He feels that at the time, "it was much less clear it would work. The noise rates were intimidatingly large and progress to improve them was very slow. Actually, the rate at which noise was being reduced was probably the same rate it is today. It's just that we've seen 20 years of steady progress, which has made everybody more optimistic." However, as the experiments got better, interest in theory increased as well. "So, what shifted was that the noise just steadily got lower and lower. The experiments just got better

and better. As a result, there was more interest from industry and from government funding agencies. And very recently, a lot more academic jobs. And so that's meant that there has been more interest in things like theory."

Professor Harrow claims that he is lucky in terms of the time that he joined the field. He believes it has grown significantly since he first joined. And with this growth, he believes that quantum information has "emerged as a topic in its own right…It used to be that quantum computing conferences were full of people who were really physicists or computer scientists or mathematicians. In that, they got their training in some discipline and then came into the field. And now there are more and more people like me, where from the very beginning we were trained in quantum computing." Harrow notes that this transition has both pros and cons. It has been beneficial in the sense that everyone is "on the same page." However, he finds the loss of "diversity of intellectual tradition" to be a shame.

What are the big challenges in quantum information?

Professor Harrow believes that the field currently faces a lot of big challenges. In particular, he feels that "theorists can be a bit wasteful in terms of the number of qubits [they] ask for." It is important that theorists find ways of doing things like error correction and fault-tolerance, while making

fewer demands of experimentalists. However, Harrow believes that the biggest challenge is actually that "classical computers are very good." This means that quantum computing researchers need to find ways to compete with classical computers for useful problems. Part of the challenge in this is that we do not actually know how to prove that certain operations will work on classical computers. In particular, Professor Harrow discussed an example of gradient descent on functions that are not convex. These functions have several local minima and, although gradient descent is not guaranteed to give a good answer, in practice it often does. "In quantum computing, because we have not been able to test our algorithms, we try to prove things, but you can often prove a lot less than you could just have a guess for and test. So I think in algorithm design, we have to figure out more heuristics." Generally speaking, Harrow believes we do not yet have a great sense of how to use quantum computers. He notes that we solely have a few ideas of how to use them for applications, like simulating molecules. Thus, he believes that a lot of work remains in coming up with better algorithms and improving the performance of currently existing algorithms.

Professor Harrow also believes that a lot of interesting questions remain in the domain of "applying quantum information as a lens to the rest of physics." He notes that people have already begun to pursue such work in the domains of the black hole information problem, quantum phase transitions, and topological order in many-body systems. "In general, there's a lot of promise for using quantum information to think about other topics in physics and I think there's a lot more to be done there."

What question is Professor Harrow currently looking to solve?

One problem that Professor Harrow is currently interested in, is figuring out how a small quantum computer could be helpful for a practical machine learning or optimization problem. In fact, he has proposed a strategy for how a quantum computer can meaningfully interact with a database that is far too large for a quantum computer to read. The solution, as it turns out, is working together with a classical computer. Thus, Harrow proposes a hybrid algorithm, in which a quantum and classical computer are used together, each doing something that the other cannot do. He claims to have a few ideas and algorithms under works, but believes there is still a lot of room for progress. "Of course, we don't understand classical machine learning fully. We just have a collection of algorithms and we don't have a general sense of how well they work. We've just seen them work on a bunch of examples and we start to gain some confidence. So, I think it's too early to hope to have one solution for all the quantum cases, but I think that would be an important area for more progress."

Harrow claims that if we want quantum computers to be useful, we "have to go where the hard computing problems are. Otherwise, why bother with the effort to make a quantum computer?" He notes that we understand a lot of these problems, such as code breaking, pretty well. There are other areas, such as quantum simulation, that we do not yet fully understand, but we have made a lot of progress on. However, he believes the domain in which we currently spend a lot of computing resources is optimization and machine learning. "I think there's just a ton of uncertainty as to how useful a quantum computer would be in this space. And so, it's not like there's one big result I would hope for that would solve it, but I think there's opportunity for a lot of progress, both in terms of theoretical progress (coming up with new algorithmic ideas and frameworks) and also really concrete things (like saying for this dataset here's an algorithm we could try and here's some idea of why we expect it to work better)." Harrow is hopeful that he could even try out approaches to these problems on a near-term quantum computer or simulate them on classical computers. This, he believes, would allow researchers to demonstrate that, given a quantum computer, they could do a much better job than solely with access to a classical device. "So, that's something where there's room for a lot of progress and I think it would be useful to make that progress." ▪

Synthetic biology can be used to create new biological systems and usually involves DNA synthesis and genetic engineering

Synthetic Biology at MIT

By Maisha M. Prome

Biotechnology is the future, and synthetic biology is just one of its sectors that are contributing to the vast quantities of research and innovation that make it just that. Using biology as a toolbox, the field treats genes as basic building blocks that can be edited, engineered, and arranged in a circuit to make cells behave in certain ways and build biological machines. From engineering yeast to make pharmaceutical compounds to creating biological computers helping astronauts survive in space, the aims and achievements of the field are as astonishing as they are important to furthering scientific discovery and saving lives. And a lot of this exciting research is happening here at MIT.

MIT has always been home to labs whose research have helped lay the foundations of the synthetic biology field. In recent years, teams of researchers, many of whom are interdepartmental, have continued to produce research and innovative applications. In 2017, "living tattoos" were developed by researchers led by Professor Xuanhe Zhao and Professor Timothy Lu. They developed a 3D printing technique using "living ink" — a hydrogel that contains living cells. These bacteri-

> *"These technologies bring us a step closer to making long missions to Mars possible."*

al cells were engineered to light up in response to different chemicals. The team printed the hydrogel into a tree pattern and placed it on a test subject's hand onto which different chemicals had been spread. They found that different parts of the tree lit up in response to the different chemicals accordingly. The cells could also be programmed to respond to signals from other cells in the hydrogel, therefore establishing the basis of a wearable microchip. The living tattoo is just one application: such a material can be printed through a nozzle into various shapes for use in further applications such as biosensors, drug release capsules, and surgical implants.

In fact, 3D printing with biological materials can have applications

that extend all the way to space. The MIT Lincoln Laboratory collaborated with the Universities Space Research Association (USRA) and NASA to write an opinion piece that was published in *Trends in Biotechnology* suggesting ideas on how DIY (Do-It-Yourself) Biology could help astronauts during long-term missions. One proposal is to engineer cells to produce biological ink as an alternative to plastics that the astronauts could use to 3D print replacement hardware. As astronauts already rely on 3D printing for these necessary parts, the proposal tackles the need for exhaustible petroleum-based plastics. Scientists on earth can engineer and test bacteria to produce certain biomaterials, and then send both the genetic engineering instructions and the 3D-printing instructions to the astronauts, who can easily replicate them to create the required part. Once used, parts can be digested and recycled into more feedstock. Similarly, the bacteria can be engineered to produce food and pharmaceuticals that are difficult to transport or store. These technologies bring us a step closer to making long missions to Mars possible.

Another team at the MIT Lincoln Laboratory wanted to address an issue relevant to all bioengineers. Engineers in other fields use multimeters, a go-to tool that allows them to measure different performance factors of their circuits. Bioengineers, however, do not have any such tool for biological systems. The team therefore aimed to create a "biomultimeter". They designed a system called PERSIA that allows

different biological functions such as transcription, translation, and enzyme functions to be monitored using fluorescent tags. PERSIA, standing for PURExpress-ReAsH-Spinach In-Vitro Analysis, uses the two elements 'Spinach' and 'ReAsh'

The MIT team at the iGEM 2019 Giant Jamboree

which fluoresce green for transcription and red for translation respectively. This allows for concentrations of different gene products to be known quickly, making the testing process of genetic circuits much faster, and therefore speeding up the timeline for research and scientific discovery.

"She decided to pursue a Ph.D. in bioengineering after realizing that brake pads in cars could be modeled after myocardial tissue in the heart."

There is also a lot of exciting research at the Synthetic Biology Center at MIT, being conducted by researchers who are passionate

about the field and come from diverse backgrounds. Dr. Shiva Razavi is a post-doctoral fellow at the Weiss Lab of the Synthetic Biology Center. Originally a mechanical engineer in the automotive industry, she decided to pursue a Ph.D. in bioengineering after realizing that brake pads in cars could be modeled after myocardial tissue in the heart. "I had no idea biological sciences could be this mathematically framed." says Dr. Razavi. Having no previous training in biological research made the transition difficult, but she began working at an immunology lab at Harvard Medical School, run by a mathematician turned biophysicist. It was her entry into the world of biological research and later synthetic biology. Dr. Razavi's current research focuses on programming stem cells to differentiate into specific cell types to give rise to different or-

gans, specifically pattern formation in the liver. Her favorite part of the research process is the problem-solving aspect and unlike many researchers, she finds the repetitive nature of experiments to be therapeutic. Each repetition uncovers something information and allows her to do hands-on work while thinking about her system at the same time, which she finds very satisfying.

Dr. Nika Shakiba also did not come from a biology background originally. As an engineering science major, it wasn't until her final year as an undergraduate that she discovered a stem cell biology lab at the University of Toronto. Excited by their research, she joined the lab to pursue a graduate degree. "As engineers, I realized throughout my Ph.D. how much we have to offer biology, and that we have a niche, that we're needed, we're wanted. That really flipped my perspective." she says. Dr. Shakiba first learned about synthetic biology on a flight on her way to a conference. Flipping through the airplane movie catalog, she found a documentary series on synthetic biology featuring MIT professors Jim Collins and Ron Weiss, who she didn't recognize at the time. Dr. Shakiba decided to join the Weiss Lab and currently works on reprogramming cells into pluripotent cells, which can then be changed into any cell type with the addition of differentiation factors. Different dosages of these factors might increase the efficiency and quality of the reprogramming process.

"Dr. Lonzaric compares his interest in synthetic biology to his interest in languages"

"The field of synthetic biology is very much in it's infancy, compared to other fields which have been around for hundreds of years." says Dr. Shakiba. The mystery of how much there is yet to know and the satisfaction of building or engineering with biology is what excites her most. Some of her favorite research in the field includes another Weiss Lab project developing cells to detect cancer in its neighbors or even themselves. Another exciting technology is barcoding cells using the MEMOIR or GESTALT systems – in which cells keep the unique DNA sequences, similar to an ID tag. When the cell divides, its subsequent generations inherit the same sequence, allowing generations to be tracked. Barcoding systems can be made smarter by including enzymes that mutate the DNA sequence for each generation on the "barcode scratchpad" so that entire family trees can be mapped out.

Dr. Jan Lonzaric is another post-doctoral fellow at the Weiss Lab who first encountered synthetic biology as an undergraduate student in Ljubljana, Slovenia through iGEM, an international synthetic biology competition. His team won the Best Project in the Health and Medicine track of the competition for their project: developing a vaccine against H. pylori, the bacteria

"What started at MIT as an IAP class in 2003 expanded to a massive international competition in only a few years."

that causes gastric ulcers. Later Dr. Lonzaric explored Transcription activator-like effectors (TALEs) for his PhD while continuing to mentor subsequent iGEM teams. He currently works on replicons, self-replicating RNAs that can be potentially used for vaccination or cancer immunotherapy, in collaboration with the Irvine Lab. Dr. Lonzaric compares his interest in synthetic biology to his interest in languages, which seem completely unrelated at first. However, they are similar in that they involve understanding of basic parts, understanding the rules that govern them as well as the exceptions that are in place, and knowing how to play around with them.

Not only are graduate and post-doctoral researchers making strides in synthetic biology. Undergraduates and high-schoolers too are contributing to the fast-growing field, especially through iGEM, just as Dr. Lonzaric did. Short for International Genetically Engineered Machine, the global synthetic biology competition is for high school students, undergraduates, and overgraduates alike. What started at MIT as an IAP class in 2003 expanded to a massive international competition in only a few years. This year alone, a total of 353 teams have participated from over 40 different countries!

Stepping into the Hynes Convention Center on the first day of the Giant Jamboree is quite an experience. The air buzzes as thousands

of Synthetic Biology enthusiasts crowd the space — participants as well as researchers and industry representatives. Teams stand out from the masses in matching T-shirts or hoodies. Hundreds of team banners line the hallways that lead to the large conference rooms where teams present to judges the projects they've worked on over the past year. Another giant room is filled with rows of easels holding up posters that display projects. Each project idea is different from the last, and span from creating synthetic milk to treating chronic kidney disease. As one participant noted "If you search up a crazy idea that could relate to engineering biology and add iGEM to it, you might really find a project out there for it." Moreover, these are solutions that tackle actual problems relating to health, medicine, environment and research.

This year, the MIT team won a gold medal for their project *The Perfect Swarm*. It is a foundational project that delves into cell-cell communication to create swarming behavior in cells. The team aimed to engineer leader cells that secrete chemokines to attract follower cells and effectively create and control cell swarms. Researchers could potentially build on this tool for further research in immunology and cell motility. Previous projects MIT iGEM teams worked on have also won awards and include diagnosing endometriosis and engineering cells in the mouth to eliminate the need for brushing teeth.

For participants, iGEM is a fun and valuable learning experience, and not just for the biology involved. For the MIT team particularly, iGEM was a taste of the entire research pipeline from start to finish. With guidance from their postdoctoral mentors at the Weiss Lab in the MIT Synthetic Biology Center, the team started their project

> *"iGEM offers more than just learning about synthetic biology. It is about real-life research and the non-science aspects that are involved"*

early in the spring, brainstorming research ideas and exploring the literature to hone in on a specific project. Over the summer, they learned the necessary laboratory skills, designed and ran experiments, and collected and analyzed results. They also held several public outreach events to engage the public regarding their project, hosted a regional

> *"Previous projects MIT iGEM teams worked on have also won awards and include diagnosing endometriosis and engineering cells to eliminate the need for brushing teeth"*

iGEM gathering, and used feedback from researchers and experts in the field to fine-tune their experiment designs and goals. The team made sure to account for bioethics in their project and also developed an impressive model of their cell swarming system.

"To me, this experience forever changed how I saw engineering." says Melody Wu, a member of the team. She has participated twice now — once in high school working on DNA supercoiling, and a second time this year at MIT. "Not only creating things at the macroscale, but also looking at the microscale. And throughout the research process, you realize how many not-so-glorious experiences you have before reaching results that may or may not have any significance. However, as part of the iGEM community you realize you are not alone in those troubleshooting experiences and those failures and successes."

iGEM offers more than just learning about synthetic biology. It is about real-life research and the non-science aspects that are involved: learning to run teams, organize projects, maintain deadlines, to communicate and collaborate with teammates – all of which are just as important as the actual science is. The Giant Jamboree itself is a giant celebration of all things biotech and genetic engineering. Just looking at the sheer number of participants, the range of projects, and the fervor with which the next generation of synthetic biologists are tackling global issues, it is easy to see that biotech is indeed our future. ∎

MURJ
UROP
Summaries

Mapping Local Government Transparency in the US

Melissa Carleton[1], Eliza Riley[2], F. Daniel Hidalgo[3]

1 Student Contributor, Wellesley College Department of Economics, Wellesley, MA 02481

2 Direct Supervisor, MIT Department of Politcal Science, Cambridge, MA 02139

3 Principal Investigator, MIT Department of Political Science, Cambridge, MA 02139

1. Summary

At MIT GOV/LAB, we are investigating what types of pressures and incentives can induce local governments in the US to improve their level of transparency. Why is government transparency important? A high level of transparency is associated with a high level of trust constituents have in their elected officials. Unfortunately, many local governments have failed to provide citizens with accessible information on their meetings, reports, and budgets. According to one study, 85% of New York county government websites failed to earn passing grades for the presence of key transparency indicators. The most common shortcoming among all websites was a lack of financial information, including budgets, taxes and fees, and expenditure (Hoefer, Curto & Savidge, 2014). A few similar initiatives have been conducted to assign report cards to other local governments across the US. However, measuring government transparency can be costly and time-intensive, making it difficult to assess on a large scale. This project uses an innovative machine learning algorithm built from scratch, which allows us to quickly measure levels of transparency for numerous local government websites.

Our algorithm scrapes websites for the presence of six transparency metrics: meeting agendas, meeting minutes, public records information, public bids, budgets, and comprehensive audited financial reports (CAFRs). Local government websites are graded on a scale of 1-6 based on the presence of these six indicators. We then generate report cards for each town displaying that town's score, how it compares to nearby towns, and how it compares to towns with similar populations. Figure 1 displays the score for a town receiving an "A" grade with all six indicators, while Figure 2 displays the score for a town receiving a "C+" grade with only three indicators. These report cards also contain descriptions of our project, the importance of government transparency, and each of our transparency metrics. Our preliminary results from a sample of over 9,000 municipalities across nearly all 50 states show that almost 70% of sites have posted up-to-date meeting agendas and meeting minutes and just over 50% of townships made their most recent fiscal year budget available; however, more than half of websites failed to post instructions for making public requests, information about bids, or finance reports (Figure 3).

Figure 1. Sample indicator summary for the town of Acworth, Georgia, which received an "A" grade (Hidalgo, 2019).

Figure 2. Sample indicator summary for the town of Dallas, Georgia, which received a "C" grade (Hidalgo, 2019).
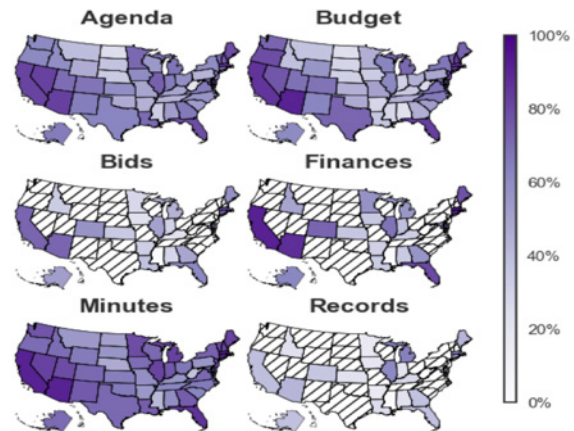

Figure 3. Proportion of municipality websites in each state containing each information type. For each map, crossed-hatched white states have yet to be evaluated for the corresponding information type (Hidalgo and Barari, 2018).

We are currently piloting a study in Georgia representative sample of local governments by sending each Georgia town's generated report card to its respective local government official. Variation in transparency metrics of over 200 towns in Georgia allows us to evaluate the causal effect of targeted interventions. Later, we will evaluate these sites to investigate whether their transparency scores improved after sending the report cards. Our approach relies on targeted "nudges" designed to incentivize local governments to improve their transparency based on behavioral design principles. According to Iris Bohnet's book What Works: Gender Equality by Design (2016), energy efficiency companies have successfully helped households reduce energy consumption
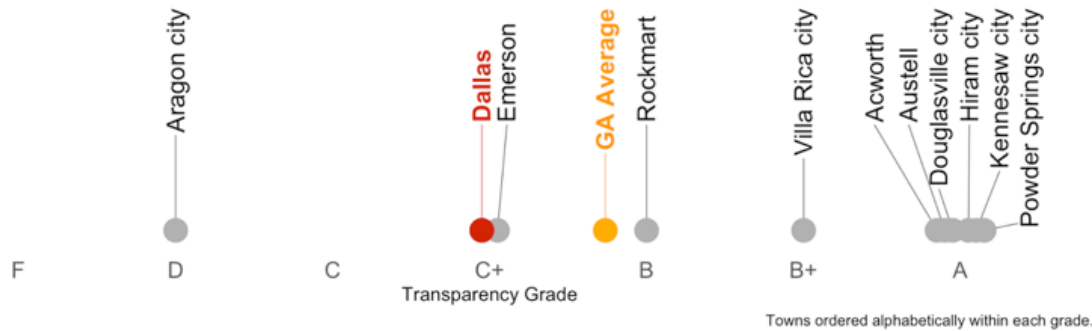
Figure 4. Graphic from the report card for the town of Dallas, Georgia displaying how its transparency score compares with nearby Georgia towns (Hidalgo, 2019).
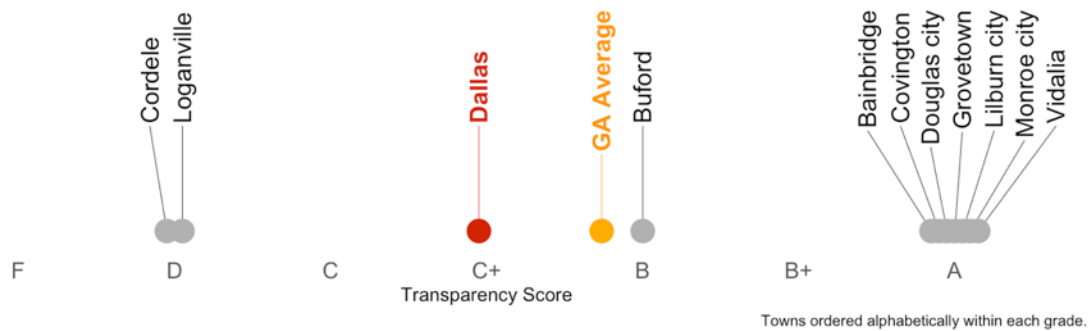

Figure 5. Graphic from the report card for the town of Dallas, Georgia displaying how its transparency score compares with similar-size Georgia towns (Hidalgo, 2019).

by providing residents reports on how their energy consumption stacks up against their neighbors. This creates "norms" such that households that overconsume energy will feel pressure to perform better when compared side-by-side to their neighbors. As mentioned, our report cards similarly display how each town compares to nearby and similar-size towns, which should also establish "norms" that incentivize underperforming governments to increase their transparency scores (Figures 4 and 5). Due to this strategic design, we predict that average transparency scores across Georgia towns will increase after local officials are sent report cards.

We are currently in the process of sending report cards to local Georgia towns. Later, we will gather data on transparency scores to see if they improved. After collecting this data, we will evaluate and refine our approach before developing a general model that explains what targeted interventions can improve local government transparency. We ultimately hope to extend this model to efficiently evaluate levels of local government transparency throughout the U.S. and, potentially, the world.

2. References
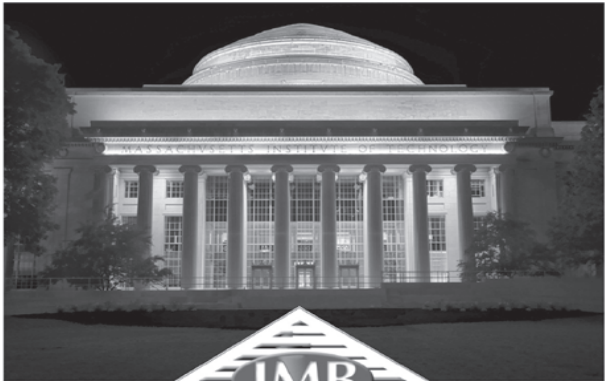
[1] Bohnet, I. (2018). *What works: gender equality by design.* Cambridge, MA: The Belknap Press of Harvard University Press.

[2] Hidalgo, F. D., & Barari, S. (2018, July). Mapping Local Government Transparency in the U.S. Retrieved from https://mitgovlab.org/updates/mapping-local-government-transparency-in-the-us/.

[3] Hoefer, T., Curto, A., & Savidge, T. (2014, October). Local Government Website Report Card 2014. *Empire Center*. Retrieved from https://www.empirecenter.org/publications/2014localgovreportcard/.

ALEXION®

JESSE
LIVING WITH gMG

**Rare Inspiration.
Changing Lives.**

At Alexion, our mission is to transform the lives of people living with rare diseases through the development and delivery of innovative medicines, as well as through supportive technologies and healthcare services. We invest in and value people who believe in the importance of our purpose and understand what it takes to deliver on it. Our passion drives us to continuously innovate and create meaningful value in all we do. In doing so, we change lives for the better – ours, people living with rare diseases and the communities we serve. Every day. **alexion.com/careers**

MURJ Reports

The Directed Migration of Neutrophil-Like Cells Through Engineered Chemokine Secretion

Yi Ran Li[1], Melody Wu[2], Margaret Zhang[3], Ron Weiss[4]

1 Student Contributor, Wellesley College, Wellesley, MA 02481

2 Student Contributor, MIT Department of Biological Engineering, Cambridge, MA 02139

3 Student Contributor, MIT Department of Biological Engineering, Cambridge, MA 02139

4 Principal Investigator, MIT Synthetic Biology Center, Cambridge, MA 02139

CELL COORDINATION WITHIN A POPULATION DEPENDS ON THE ABILITY OF AN INDIVIDUAL CELL TO ACCURATELY RECEIVE AND RESPOND TO EXTRACELLULAR STIMULI FROM THE ENVIRONMENT AND NEIGHBORING CELLS. COORDINATING CELLULAR MOTILITY, WHERE CELLS MOVE IN RESPONSE TO EXTERNAL CUES, IS CENTRAL TO MANY PHYSIOLOGICAL RESPONSES. THE MIGRATORY BEHAVIOR OF NEUTROPHILS TOWARDS THE HIGH CONCENTRATION REGION ALONG A CHEMOKINE GRADIENT IN THE HUMAN ADAPTIVE IMMUNE SYSTEM IS A PRIME EXAMPLE OF THIS. HERE, WE PRESENT A MECHANISM TO HARNESS CELLULAR CHEMOTAXIS TO CONTROL CELLULAR SWARMING AND DIRECTED MOVEMENT. WE ENGINEERED HUMAN EMBRYONIC KIDNEY (HEK) CELLS TO SECRETE CHEMOKINES IN ORDER TO INDUCE CHEMOTAXIS IN WILD-TYPE NEUTROPHIL-LIKE HL60S, A MODEL FOR ENDOGENOUS NEUTROPHILS IN THE BODY. TO EVALUATE CHEMOTAXIS, WE FIRST DIFFERENTIATED HL-60 CELLS INTO CHEMOTACTIC NEUTROPHILS. WE THEN INTRODUCED CHEMOKINES PRODUCED BY OUR ENGINEERED HEK CELLS AND EVALUATED NEUTROPHIL MOVEMENT UTILIZING SEVERAL CELLULAR MIGRATION ASSAYS. WE ANTICIPATE OUR ENGINEERED SYSTEM WILL PROVIDE INSIGHT INTO HOW IMMUNE SYSTEMS DEVELOP AS WELL AS FORM A PRELIMINARY TOOLBOX FOR RECRUITING MAMMALIAN CELLS SELECTIVELY IN TISSUE ENGINEERING APPLICATIONS.

1. Introduction

Swarm behavior is defined as a large group of individuals performing a collective function, usually movement, without every member directly being aware of the overall goal. The leaders determine the swarm's motion by influencing the rest of the group, the followers. These basic leader-follower interactions allow for more complex dynamics and organized function by the swarm. Swarming behavior can be observed in nature among birds and insects, and have been recreated using robots (Rubenstein 2014). Swarm behavior has also been observed at the microscopic level among cells through the process of chemotaxis.
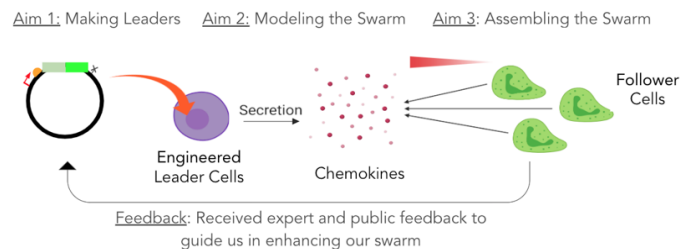

Figure 1. Outlining our experiment process according to the above three aims with feedback integrated into our research.
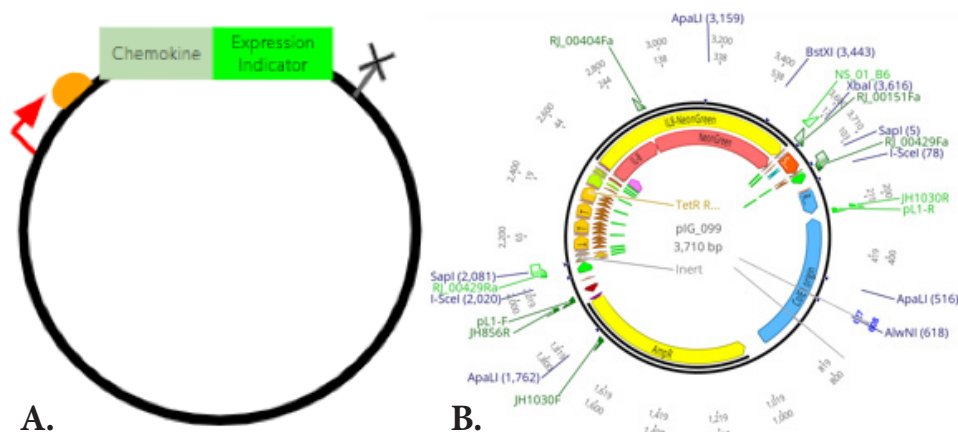

Figure 2a) The collection of plasmids genetic circuits we cloned had this general structure of promoter, kozak sequence (translation initiation site), chemokine gene with an optional fused expression indicator (such as NeonGreen), a synthPolyA terminator, and a vector with Ampicillin/Carbenicillin resistance. **b)** Example of chemokine plasmid with IL8-NeonGreen genetic circuit with Ampicillin resistance. These plasmids were assembled through a process called Modular Cloning Golden Gate Assembly, which allows you to put various parts (i.e. promoter, kozak, coding sequence/gene, terminator, etc.) together to create complete genetic circuits, similar to a typical electric circuit.
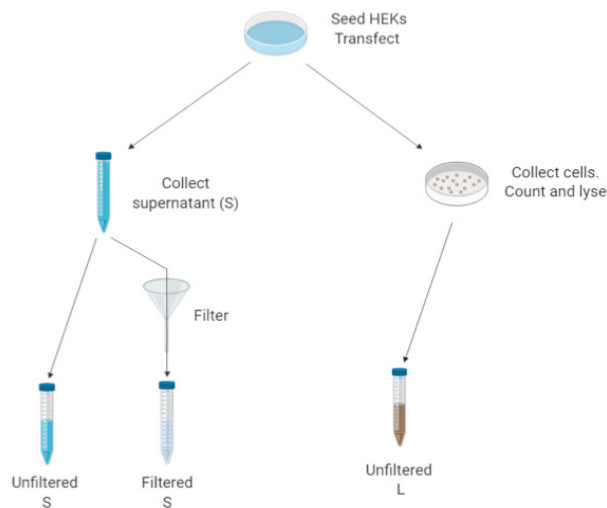
Figure 3. Secretion test outlined in diagram shown. Supernatant was collected and concentrated through filtration to better measure expression. After the supernatant was collected, the cells remaining were collected and lysed. These samples were then run through a plate reader and measured for fluorescence.

Chemotaxis is the directed movement of cells up or down a chemical gradient, and plays a critical role in a variety of biological systems such as in the immune response, cancer metastasis, and the development of multicellular organisms (Legler 2018). The chemical stimulus, known as a chemokine, is secreted by certain cells and causes other cells to polarize toward or away from the chemokine source, resulting in chemotaxis. The chemokine is a ligand activating signaling molecules such as GPCR receptors upstream of the actin polymerization pathway. The swarming nature of certain white blood cells in the human body's natural adaptive immune response is currently the most widely used model for chemotaxis. When dealing with infections, neutrophils, the largest group of white blood cells, detect certain chemicals released by pathogens and subsequently follow, surround, and destroy the pathogens (Petri 2018). Cells in damaged tissues can release chemokines to attract neutrophils to the area and stimulate the repair process.

In this proof-of-concept, we engineered HEK cells to secrete chemokines that would induce chemotaxis in neutrophils. We evaluated neutrophil movement towards secreted chemokines through a series of cellular migration assays. In order to evaluate the experimental parameters that would produce meaningful results as well as represent real-world conditions, a computational model
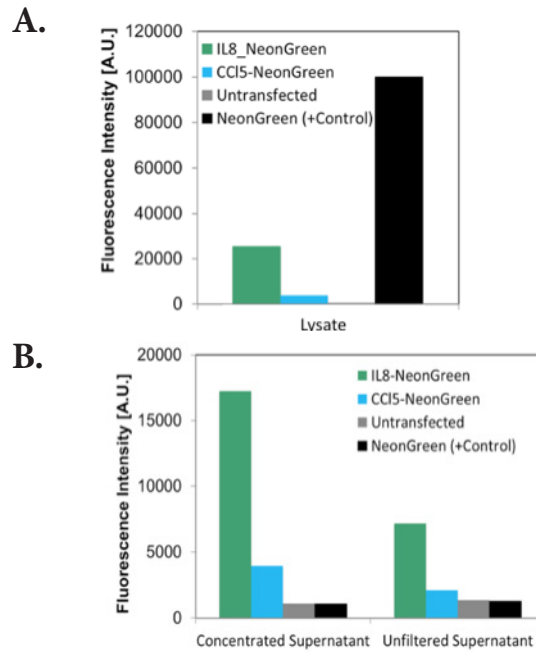

Figure 4. Secretion test results shown. **a)** In the lysate, which is representative of expression within the cells, there is presence of IL8-NeonGreen (green) and CCL5-NeonGreen (blue). When comparing to NeonGreen positive control, it is expected to be lower as unlike the positive control, the chemokines with fused NeonGreen are secreted outside of the cell. We also note that the CCL5-NeonGreen transfection was not as efficient as for IL8-NeonGreen. **b)** In the supernatant, which is representative of all that is secreted outside of the cell, there are clear presence of IL8-NeonGreen (green) and CCL5-NeonGreen (blue). Comparing this with the negative and positive controls which have low fluorescence intensity that is likely due to phenol red in the culture medium.

using the software Morpheus was developed to generate various simulations that would aid us in our experimental assays. We anticipate that our engineered system will provide insight into cell-cell communication during the human body's immune response and form a preliminary toolbox for recruiting mammalian cells in various tissue engineering applications.

2. Methods

As part of iGEM 2019, our team designed our project with the following methodology and goals in mind to synthesize swarming behavior: 1) making leaders, 2) modeling signaling elements integral to swarm-like motion, and 3) assembling the swarm (Figure 1).
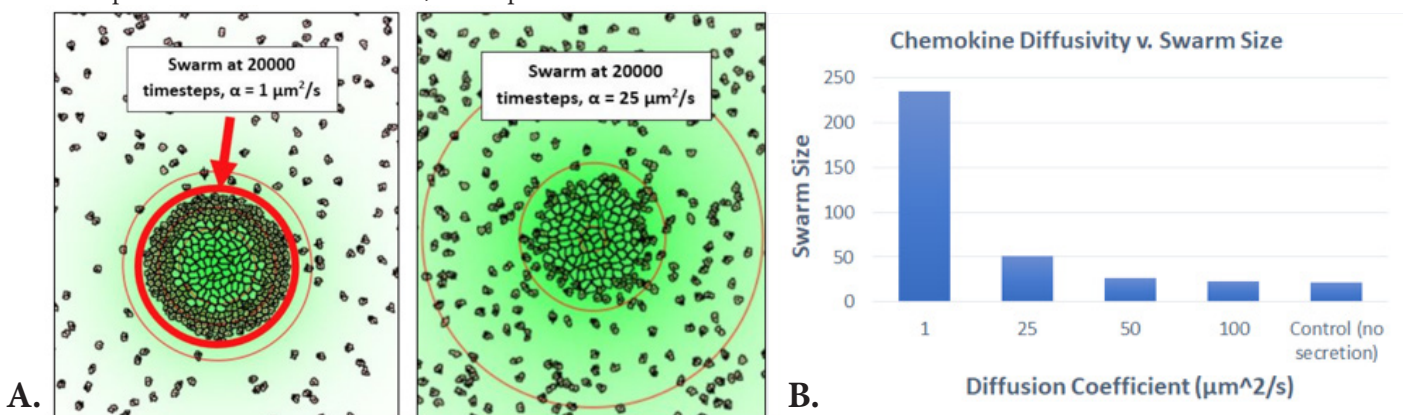

Figure 5a) Cell swarming at 20000 timesteps, with a diffusion coefficient of 1 μm^2/s on the left and 25 μm^2/s on the right. On the left, an example of a cell swarm is pointed out. All Follower cells (gray) in the swarm are adjacent to other Follower cells in the swarm or to Leader cells. **b)** Swarm size was determined by manually counting the number of Follower cells within the group at the center.
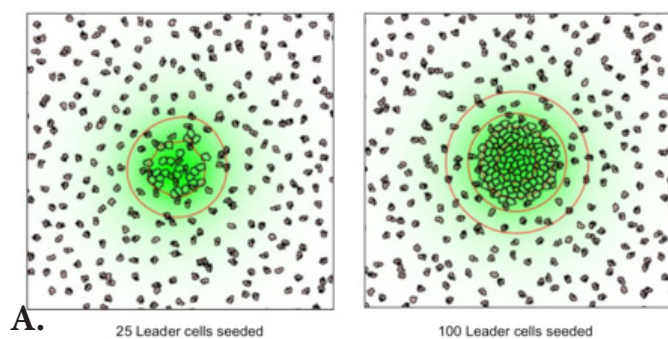
Figure 6a) Cell swarming at 20000 timesteps with a diffusion coefficient of 25 µm^2/s and 25 Leader cells seeded on the left and 100 on the right. **b)** Follower cells in the swarm were counted to determine swarm size.
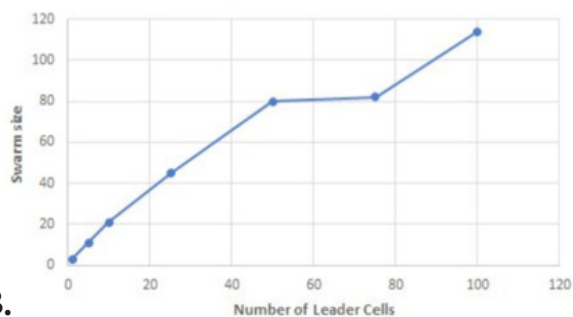
As a proof-of-concept, we focused on a system involving HEK-293 cells, which served as our engineered leaders, and HL-60 cells, which served as our followers. HEK-293 cells were chosen because they are a human cell line, they are adherent to surfaces so we can centralize them to a specific location, and because they have high transfection efficiency. HL-60 cells were chosen because they are shown to exhibit neutrophil-like chemotactic behavior following differentiation in 1.3% DMSO (Fleck 2005). We envision that this system could be used in the future for other leader-follower setups to further conduct research in immunology. Making leaders first involved designing our genetic circuits and then cloning them with the appropriate mammalian genetic components and indicators (design in Figure 2a, example in 2b) through Golden Gate Modular Cloning Assembly.

This cloning toolbox was also created for the purposes of future iGEM teams or researchers interested in studying immunology. HEK cells were then tested for viability in HL60 media, RPMI, for later assays that involved both cell types. Following validation that the HEKs were transfectable and viable in RPMI (the media for HL60 culturing), we performed chemokine secretion test assay by the HEK leader cells (process outlined in diagram below). We wanted to assess whether engineered HEKs could produce and secrete the IL8 and CCL5 chemokine, which was fused to a NeonGreen reporter for easy readout. To verify this, the supernatant of engineered HEKs expressing IL8-NeonGreen was isolated, which includes any protein that was secreted by the cell, as well as the lysate, which included anything expressed but not released by the cell.

Notable expression of NeonGreen in both our positive control and also cells engineered to express IL8-NeonGreen and CCL5-NeonGreen by our leader HEK cells (Figure 4a) was observed, indicating that the HEKs could successfully express our engineered construct. Secretion of our chemokines of interest, which included IL8-NeonGreen and CCL5-NeonGreen, by our engineered HEK cells was also verified (Figure 4b). We proceeded to focus on IL8 following this experiment, as it has been shown to be one of the primary chemokines inducing neutrophil-like HL60 chemotactic behavior (Stålhammar 2016).

3. Modeling

We generated a simulation of engineered chemotaxis using the software Morpheus for directed chemotaxis and cell swarming based on the Cellular Potts Model (Graner 1992), developed by Fracois Graner and James Glazier, in order to optimize our in vitro cellular motility experiments. Stable and steep chemokine gradients are critical for cellular swarming of the differentiated HL-60 follower cells. Current methods of gradient formation involve microfluidic devices or micropipettes, but these do not factor in cells as a potential source of the chemokine. With this model, we tested parameters such as the diffusivity of the secreted protein, number of leader cells seeded, and cell placement to determine the optimal conditions and swarm size for our system. The model served as an important tool that enabled us to simulate the formation of chemokine gradients and the response of chemotactic cells to this.

To test the diffusivity of the secreted protein, an important parameter that determines how the chemokine gradient is formed, we ran a simulation with 90 seeded HEK cells in the center of a 1000 µm by 1000 µm lattice, and 900 seeded HL-60 cells throughout (Figure 5). We tested the diffusion coefficient at 1, 25, 50, and 100 µm^2/s to examine the resulting swarm size. Our control constituted the case of no IL-8 secretion, which prevented any gradient from forming. Based on our data, we observed that a higher diffusion coefficient resulted in fewer Follower cells attracted. There is effectively no swarm formation at α = 100µm^2/s, as the 'swarm size' is equal to that of the control. This informed us that swarming would not occur with the secretion of pure IL-8, which has an estimated diffusion coefficient of 200 µm^2/s, in RPMI media; in order to create a gradient steep enough for swarming to occur in our assays, we needed to decrease the diffusion coefficient of the chemokine, either by increasing the viscosity of our media by using hydrogels instead of liquid media, or increasing the size of the chemokine by adding a fluorescent tag. In order to inform us about seeding density, we tested how the amount of Leader cells at the center of the lattice affected swarm size (Figure 6). We used the same parameters as the diffusion coefficient experiment and set the diffusion coefficient to 25µm^2/s. Leader cell density was tested at 1, 5, 10, 25, 50, 75, and 100 cells.
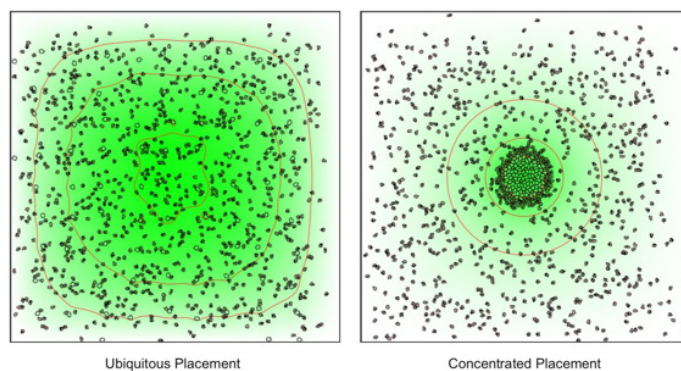

Figure 7. Cell swarming at 20000 timesteps with a diffusion coefficient of 25 µm^2/s with ubiquitous seeding of the HEK cells on the left and concentrated seeding on the right.
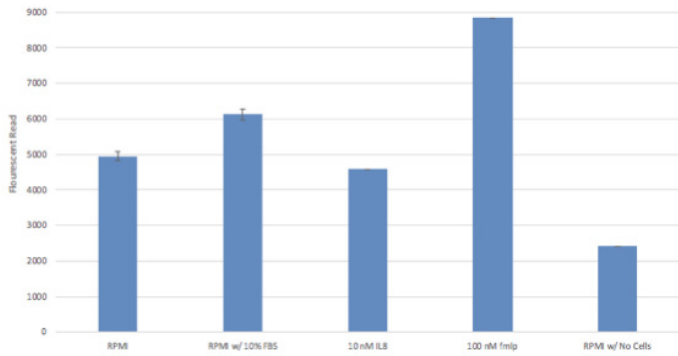
Figure 8. Microplate reading of DNA stained with CyQUANT dye shows differentiated HL-60 response to RPMI, RPMI with 10% FBS, 10nM IL8, 100 nM fMLP, and RPMI with no cells.
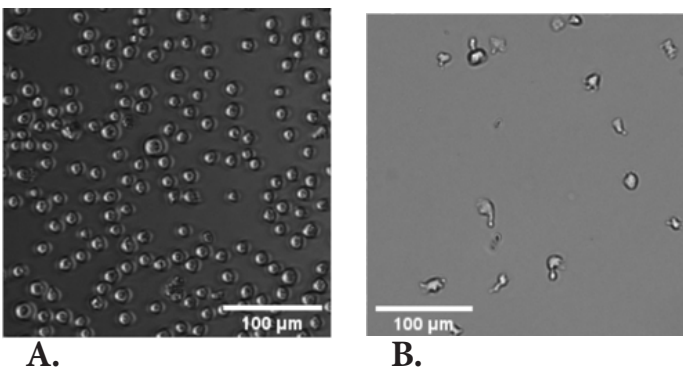

A.   **B.**

Figure 9a) Time-lapse microscopy experiment of undifferentiated cells with chemical stimuli show flow-based movement. **b)** Time-lapse microscopy experiment of differentiated cells with chemical stimuli show actin-based chemokinesis.

The swarm size was similarly manually counted. The results show that a larger swarm size is dependent on a higher seeding density. To test the placement of Leader cells on swarm formation, we simulated two different configurations of Leader cell placement, concentrated and ubiquitous with the same parameters as the prior two experiments and a diffusion coefficient of 25 μm^2/s (Figure 7). Our model predicts that swarming behavior is more apparent with a concentrated placement of Leader cells, as the gradient formed was steeper than that of ubiquitous placement. With this model, future improvements can be made in terms of design so that we might be able to engineer swarm size and movement or patterning.
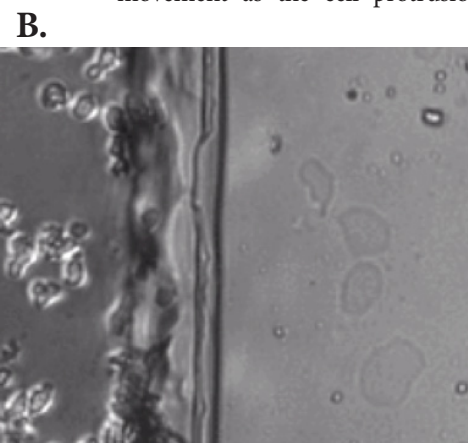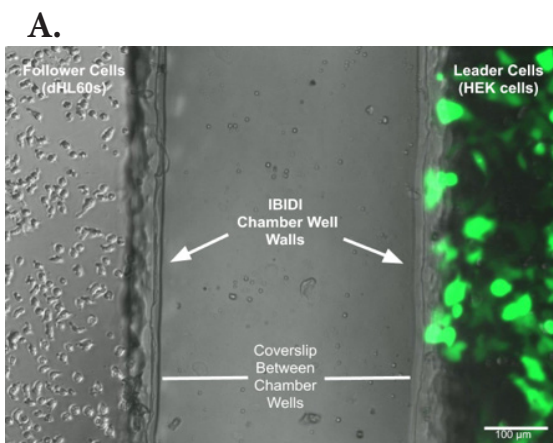
4. Preliminary Data and Results

To create cell swarms, we needed to demonstrate the ability for our cells to sense chemokines, move in their presence, and chemotaxis in response to suitable chemokine gradients. We employed Boyden chamber assays, time-lapse microscopy, and IBIDI chamber experiments to assess the assembly of our cell swarms.
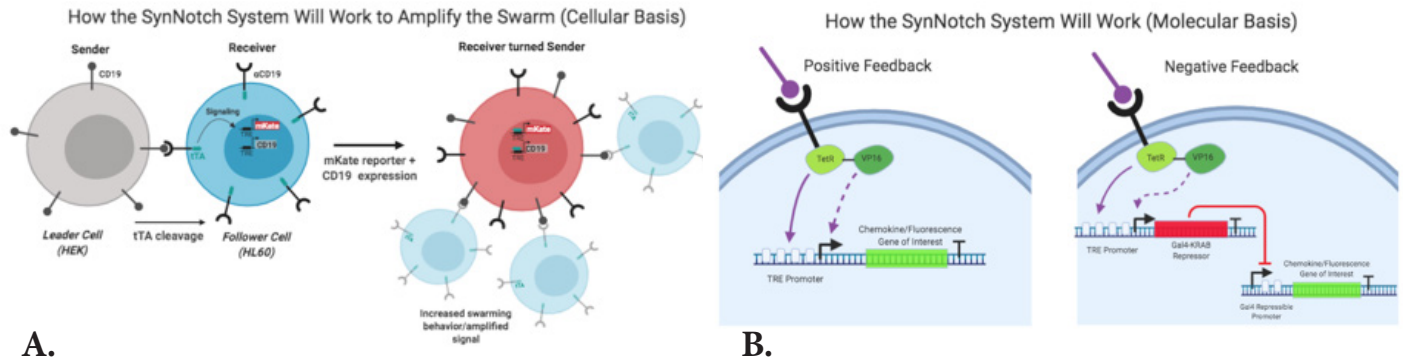
Boyden chamber assays quantify the chemotactic index, a measurement of how differentiated HL-60s respond to chemoattractants. HL-60s were put in wells above a 24 well plate containing chemokines. When the neutrophils sense chemokines beneath, they squeeze through the 3 um pores that divides the well and the plate. Figure 8 demonstrates that in the presence of 100nM fmlp, the cells moved towards the chemokine, demonstrating their ability to sense these chemicals. With the positive control, we concluded that the lack of movement in the presence of IL8 is due to error in the preparation of the chemokine.

The time-lapse experiments were run under multiple conditions, one of which was differentiation with optional stimulation. Microscopy visualizes the motility index, whether the cells move and are properly differentiated. When HL-60s were undifferentiated, movement was based on flow and no crawling behavior was observed. On the other hand, when HL-60s were differentiated, the cells crawled along the slide with noticeable leading and lagging edges likely attributed to actin buildup and breakdown.

The swarming ability combines the chemotactic and mobility index tested in the assays above. In the IBIDI chamber experiment, differentiated HL-60s and secreting HEK cells were seeded on two sides of a silicon wall. The chemokine secreted by HEK cells built up on one side of the wall, creating a sharp gradient. Neutrophils move in the path of least resistance. Hence, the movement beneath the silicon wall towards HEK cells demonstrates chemotactic behavior.

5. Conclusions

The results from the experiment demonstrates that the IL8 circuit we design was transfected and expressed by our leader cells. In addition, we were able to culture differentiated follower cells that demonstrate mobility. The HL-60s moved towards the fmlp control in the Boyden chamber and changed their morphology to crawl beneath coverslip and the IBIDI Chamber. Based on the movement direction, we can assume that they will continue to move towards leader cells and their shapes also convey rightward movement as the cell protrusion faces the leader cells. This

A.   **B.**

Figure 10a) Differentiated HL60s and secreting HEK cells are cultured on different sides of IBIDI Chamber Walls **b)** Differentiated HL60s crawl beneath IBIDI chamber due to chemokine gradient built across the silicon wall.

A.

B.

Figure 11a) The synNotch system will work to amplify the swarm through a ligand-receptor cell contact system (contrasting to our long-range diffusion based signaling with chemotaxis). Follower receiver cells will swarm to the location of the leader cells, and upon contact, an engineered or endogenous expression system will be activated amplifying chemokine signal. **b)** Two possible forms of feedback upon contact of the sender cell (indicated in purple) with the receiver cell (indicated in black) could be developed: positive or negative feedback.

indicates that we were able to create the hallmark of a swarm by creating a leader that secretes chemokines and induce directed cell movement. Our experiment demonstrates a first step towards making a toolbox for the immune system.

6. Future Directions

While results indicate the hallmark of a swarm with follower cells moving towards our leader cells, there is a need for further assay and experimental testing to solidify results by rerunning the Boyden Chamber assays with supernatant from our HEK cells and with pure IL8, setting up control assays with the IBIDI chambers, and further testing with under agarose assays to verify motility. Further work would involve looking into building a more complex responsive swarm.

Upon receiving feedback from experts, we determined that controlling swarm size and mimicking natural swarming behavior should be a large priority. We began development of a synNotch system genetic circuit collection to implement positive and negative feedback systems to control swarm size. The synNotch system was developed by the Wendell Lab at the University of California San Francisco, specializing in understanding cell-cell communication (Roybal et al. 2016). The system upregulates certain marker genes via surface receptor binding when the leader and follower cell come into contact. Future work would involve building complex systems using the synNotch method. While we were not able to test the synNotch system in an experimental setting, implementation and further research would be valuable.

7. References

[1] Fleck, R. A., Romero-Steiner, S., & Nahm, M. H. (2005). Use of HL-60 cell line to measure opsonic capacity of pneumococcal antibodies. Retrieved August 15, 2019 from https://www.ncbi.nlm.nih.gov/pubmed/15642980.

[2] Graner, F., & Glazier, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. Physical Review Letters, 69. Retrieved November 10, 2019 from https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.69.2013

[3] Heit, B., & Kubes, P. (2003). Measuring Chemotaxis and Chemokinesis: The Under-Agarose Cell Migration Assay. Retrieved August 5, 2019

[4] Legler, D. F., & Thelen, M. (2018). New insights in chemokine signaling. F1000Research, 7. Retrieved on November 10, 2019 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5782407/.

[5] M Keenan, Thomas & Folch, Albert. (2008). Biomolecular Gradients in Cell Culture Systems. Lab on a chip. 8. 34-57. 10.1039/b711887b.

[6] Park, J. S., Rhau, B., Hermann, A., et. al. (2014, April 22). Synthetic control of mammalian-cell motility by engineering chemotaxis to an orthogonal bioinert chemical signal.

[7] Petri, B., & Sanz, M.-J. (2018). Neutrophil chemotaxis. Cell and Tissue Research, 371(3), 425–436. https://doi.org/10.1007/s00441-017-2776-8

[8] Roybal, Kole T., et al. "Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors." Cell, vol. 167, no. 2, 2016, doi:10.1016/j.cell.2016.09.011.

[9] Rubenstein, M., Cornejo, A., Nagpal, R. 2014. Programmable self-assembly in a thousand-robot swarm. Science, 345, 6198.

[10] Ryman, Adam. [Adam Ryman]. (2008). White Blood Cell Chases Bacteria [Video file].

[11] Sherman, H., & Rothenberg, M. (2012). Cell Migration, Chemotaxis and Invasion Assay Using Staining. Retrieved August 5, 2019

[12] Stålhammar, M. E. (2016, August 22). Differential neutrophil chemotactic response towards IL-8 and bacterial N-formyl peptides in term newborn infants. Retrieved from https://www.tandfonline.com/doi/full/10.1080/03009734.2016.1228721.

8. Attributions

Studying Resonant Interactions and the Use of Electron Data for DUNE Energy Reconstruction

Anjali Nambrath[1], Minerba Betancourt[2], Or Hen[3]

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

2 Direct Supervisor, Fermi National Accelerator Laboratory, Batavia, IL 60510

3 Principal Investigator, MIT Department of Physics, Cambridge, MA 02139

ONE OF THE GOALS OF THE LBNF/DUNE PROJECT IS TO OBTAIN VALUES FOR NEUTRINO OSCILLATION PARAMETERS. CALCULATING THESE VALUES REQUIRES ACCURATE KNOWLEDGE OF THE INCOMING NEUTRINO BEAM ENERGY. RECONSTRUCTING THIS ENERGY FROM MEASUREMENTS OF FINAL-STATE PARTICLES IS CRITICAL TO DUNE'S SUCCESS. WE STUDY CALORIMETRIC ENERGY RECONSTRUCTION WITH DUNE QUASI-ELASTIC EVENTS SIMULATED USING GENIE, AND EXPLORE THE POSSIBILITY OF CONSTRAINING THE RECONSTRUCTION USING ELECTRON SCATTERING DATA FROM CLAS AND GENIE SIMULATIONS. WE ALSO COMPARE ELECTRON AND NEUTRINO SCATTERING IN THE RESONANT CHANNEL, WITH THE GOAL OF EXTENDING THESE RECONSTRUCTION STUDIES TO MORE TYPES OF EVENTS IN THE FUTURE.

1. Introduction

Neutrinos are among the fundamental particles of our universe. They have no charge, and they are incredibly small and very light. They are also the most abundant massive particle in the universe, but they interact incredibly rarely (only via gravity and the weak force). Three flavors of neutrinos have been discovered: the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Neutrinos oscillate, or change flavor, as the travel as a consequence of their mass (Gil-Botella, 2015). The discovery of neutrino oscillations and the realization that neutrinos are not massless, as previously thought, led to the 2015 Nobel Prize in Physics.

The study of neutrinos is key to several major outstanding questions in particle physics, and is also a gateway to physics beyond the Standard Model (Fantini et al, 2018). From testing the weak force to CP violation to various astrophysical processes, neutrinos have the potential to provide deeper insight into our universe than ever achieved before. Particularly, the phenomenon of neutrino oscillations indicates new physics, and as a consequence it has become the focus of much new research.

The goal of several upcoming neutrino experiments, including the LBNF/DUNE project at Fermilab, is to measure values for neutrino oscillation parameters such as the mixing angles and mass differences (which describe the frequency and amplitude with which the flavor change occurs) (Acciarri et al, 2016). These measurements are obtained by observing neutrino appearance and disappearance probabilities. In order to extract the parameter values from these measured probabilities, it is necessary to have an accurate measurement of the energy of the neutrino beam. This energy, however, has a high uncertainty because neutrino beams are produced via fixed target collisions and are by no means monoenergetic. The beam energy has to be reconstructed from measurements of final-state particles observed by detectors. Finding a good model for energy reconstruction is critical to the success of these experiments. In this study, we primarily look at the calorimetric approach where the energies of final-state particles are summed.

There is also another approach to the energy reconstruction. Monoenergetic electron beams can be produced to extremely high precision, and electrons and neutrinos seem to display similar scattering behavior in some interaction channels. Therefore, it is potentially possible to use electron scattering data to constrain neutrino energy reconstruction models in order to obtain better results (Ashkenazy et al, 2018). In this study, we perform a proof-of-concept test by using data from the CLAS experiment at Jefferson Lab and GENIE electron simulations to constrain neutrino energy reconstruction calculations. We also continue previous work by comparing electron and neutrino scattering in resonant interactions, with the hope of extending these studies to more types of events.

2. Methods

In this study, we simulate neutrino and electron scattering events using GENIE's Monte Carlo generator, version 3.0.4. GENIE is a software framework that implements neutrino event generation using a comprehensive physics model and tools that simulate neutrino interactions (Andreopoulos et al, 2015).
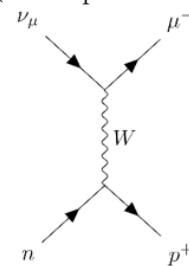

Figure 1. Neutrino scattering event with no pion production.

Event generation. The GENIE Monte Carlo generator handles electron and neutrino scattering across a full range of interaction channels and many nuclear targets. We simulated muon neutrino and electron beams at 2.2 GeV and in the range from 0 to 10 GeV, using the DUNE flux spectrum (which describes the
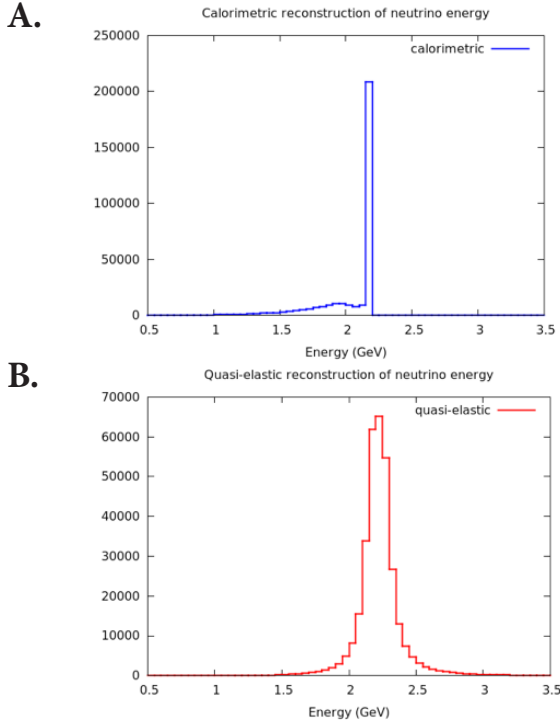
A.

B.

Figure 2. Comparing reconstruction methods in the case of monoenergetic neutrinos. **a)** Calorimetric reconstruction of 2.2 GeV neutrinos scattering on a carbon-12 target. **b)** Quasi-elastic reconstruction of 2.2 GeV neutrinos scattering on a carbon-12 target.

expected energy distributions of DUNE data). GENIE's CCQE, CCRES, Default+MEC generators were used for neutrinos, and the EMRES, EM+MEC, and Default+MEC generators were used for electrons. Events with simulated with fixed targets of carbon-12 (for comparison to CLAS data), and argon-40 (to be used in the time-projection chamber detectors at DUNE).

Event selection. In the first part of this study, we consider only resonant interactions. Later, we consider events in all channels. For the neutrino scattering resonant studies, we consider only charged current (CC) events in order for more direct comparison to the electron events -- in this way the final products of both processes include charged leptons.

For the DUNE energy reconstruction studies, we perform some cuts to obtain a purer sample. We require that the four-momentum transfer Q^2 is greater than 0.5 GeV and the square invariant mass W is less than 2.0 GeV. We also require that exactly one proton is produced with momentum greater than 0.3 GeV and that no pions are produced. For the resonant study, we cut on events with a pion momentum greater than 150 GeV.

3. Energy Reconstruction Studies

The goal of these studies was to observe the accuracy of different energy reconstruction methods in various different contexts. We considered a calorimetric approach and the quasi-elastic hypothesis, and ultimately went with the calorimetric method. Since we are performing 0pi cuts on the events, the collisions we are studying resemble those in Figure 1.

The quasi-elastic hypothesis assumes conservation of momentum during the interaction and uses kinematic arguments (Saraswat, 2017). The reconstructed energy it suggests is given by the following expression:

$$E_{rec} = \frac{m_n^2 - \left(m_p - E_b\right)^2 - m_\ell^2 + 2\left(m_p - E_b\right)E_\ell}{2\left(m_p - E_b - E_\ell + p_\ell cos\theta_\ell\right)}$$
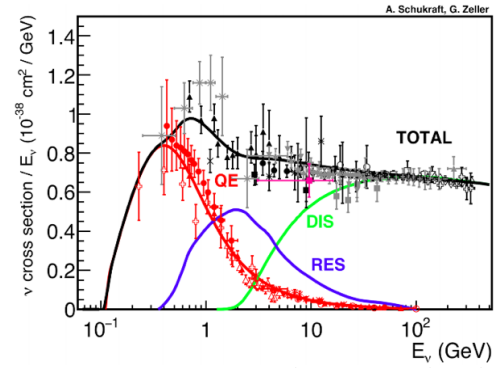

Figure 3. Neutrino scattering cross-sections by interaction channel. In the 0-10 GeV range, relevant for DUNE, all three channels are important.
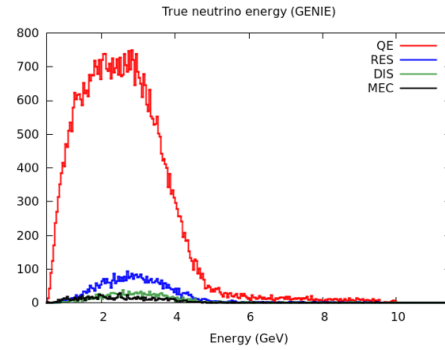

Figure 4. DUNE-like events after QE and 0pi cuts.

Eb, the binding energy of the proton, was given a value of 0.025 for carbon and 0.04 for argon.

The calorimetric approach, on the other hand, assumes conservation of energy, and the reconstructed neutrino energy is given by the following sum of the energies of the scattering collision's final products:

$$E_{rec} = E_\ell + K_p + E_b$$

Comparison using monoenergetic neutrinos. In order to observe the difference between the two reconstruction methods in the simplest possible case, we first consider monoenergetic neutrinos. Specifically, we use simulated quasi-elastic νμ scattering events with a carbon-12 target at 2.2 GeV. The resulting reconstruction plots are shown in Figure 2. In the monoenergetic case, the calorimetric method appears to be the better approach.

Comparison with DUNE flux. Having verified both reconstruction methods on 2.2 GeV data and identified the calorimetric approach as more accurate, we now analyze DUNE-like events. νμ events in all channels were generated using an argon-40 target at energies between 0 and 10 GeV with the DUNE flux spectrum. All channels were used for generation because in DUNE's operating range, neutrino scattering events in the quasi-elastic (QE), resonance (RES), and deep inelastic (DIS) channels all have comparable cross-sections (Figure 3). Despite DUNE being dominated by non-QE events, we perform a variety of cuts on this data (as listed in the Methods section) to give a mostly QE sample. Resonant and other non-QE events will be the subject of further

Table 1. NuFIT 4.0 Oscillation Parameters (normal ordering, without SK atmospheric data)

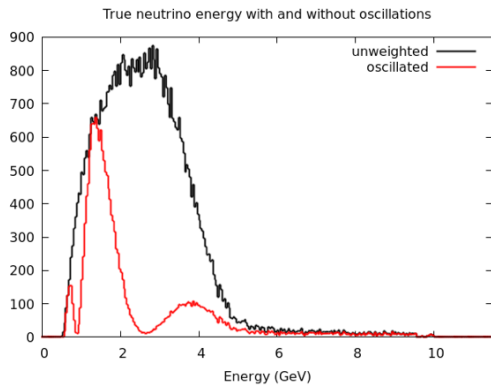|	
$sin^2 \theta_{12}$	0.310
$sin^2 \theta_{23}$	0.580
$sin^2 \theta_{13}$	0.02241
Δm_{21}^2	7.39×10^{-5} eV
Δm_{atm}^2	2.525×10^{-3} eV
δ_{CP}	0
DUNE beam dip angle	5.8°

Figure 5. Oscillated DUNE neutrino energies compared to the unweighted energy spectrum.
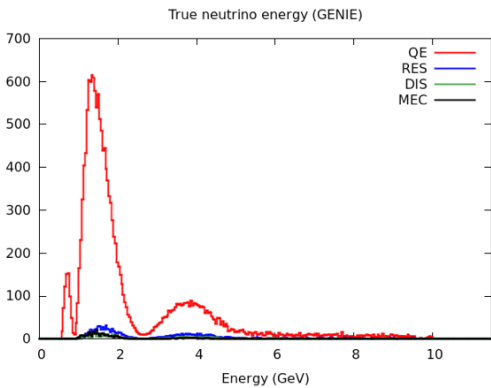

Figure 6. Oscillated DUNE neutrino energies, broken down by interaction channel.
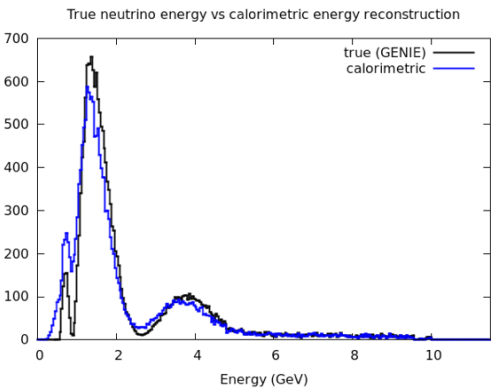

Figure 7. Calorimetric reconstruction of oscillated DUNE neutrinos..

studies. After applying the QE and 0pi cuts, we get a sample that looks like the plots in Figure 4 – clearly, we have a mostly quasi-elastic data set.

Adding neutrino oscillations. DUNE is a neutrino oscillation experiment, so including the DUNE flux is insufficient to properly simulate DUNE events. The Prob3++ package was used to weight each event by the νμ survival probability at that energy, using NuFIT 4.0 results as parameters, listed in Table 1 (Esteban et al, 2018). Figure 5 shows the energy spectrum of the oscillated neutrinos as compared to the true energy without oscillations. Figure 6 shows the oscillated energy broken down by channel: quasi-elastic (QE), resonance (RES), deep inelastic scattering (DIS), and meson exchange currents (MEC). Reconstruction gives the results in Figure 7, shown with the true energy for comparison.

Using CLAS electron data. It has been established that there is enough similarity between electron and neutrino quasi-elastic scattering events to try to constrain neutrino energy reconstruction using electron data. We perform a rough demonstration of this using CLAS fractional feed-down data to constrain our simulated
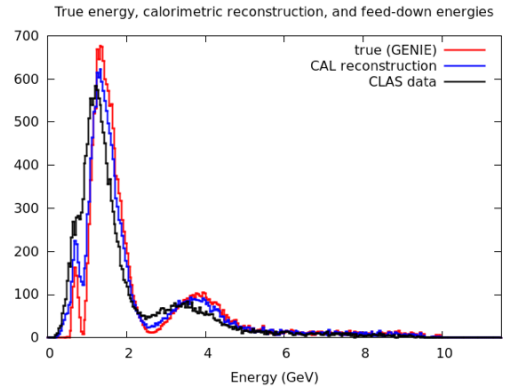

Figure 8. Feed-down energy using CLAS data, calorimetric reconstruction, and GENIE true energy with a carbon-12 target.
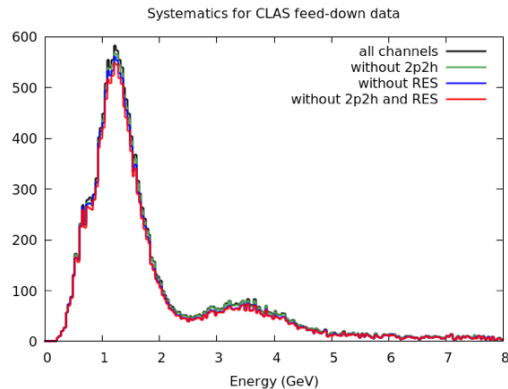

Figure 9. Feed-down energy systematics using CLAS data and a carbon-12 target. Contributions from non-QE channels result in a very small effect.

DUNE oscillation events. CLAS data exists for carbon-12 targets, so for these studies we generate νμ events in all channels with a carbon target in the range of 0 to 10 GeV with the DUNE flux applied. Plots of the oscillated energies and calorimetric reconstructions with the carbon-12 target can be seen in Figure 8.

We now apply the CLAS data to the true energies. The data we are using comes from events with a 2.2 GeV electron beam and carbon-12 targets. These energy distributions are essentially convolved with the true energies from our νμ GENIE simulations. The resulting energies with the feed-down thus applied are also shown in Figure 8. Clearly there is an effect on the results from the application of CLAS data that is not present in the straightforward calorimetric reconstruction. Systematics are shown in Figure 9 – the effect of non-QE channels on the reconstructions is quite small.

We also compare the above results, using CLAS data, to results obtained by using simulated electron events. We used electron scattering events generated with a carbon-12 target at 2.2 GeV,
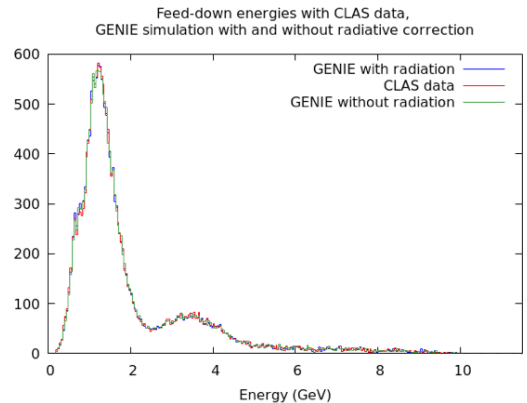

Figure 10. Feed-down energy using CLAS data, simulated data with and without radiative correction (using a carbon-12 target).
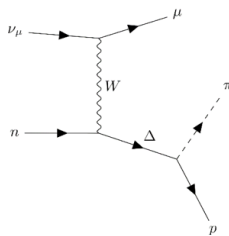
Figure 11. Resonant pion production via neutrino scattering.

with GENIE and with a different simulation that also included a radiative correction. These distributions were applied to the oscillated DUNE neutrino energy spectrum. Figure 10 shows the results with this simulated data compared to the distribution obtained using actual carbon-12 CLAS data.
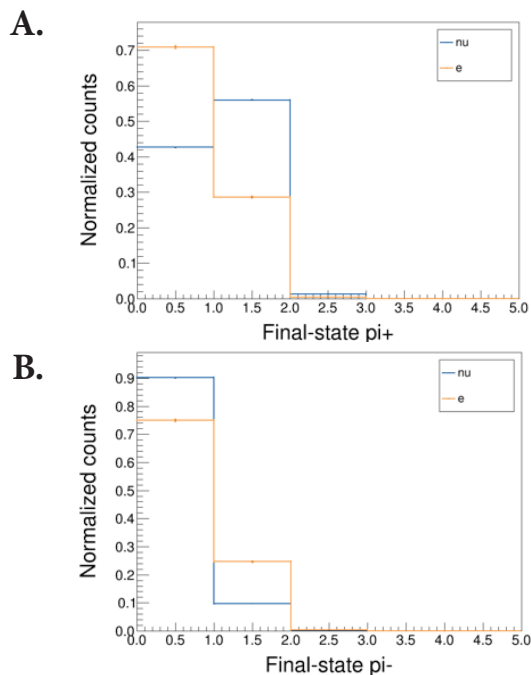

Figure 12. Fairly substantial differences in pion production between the electron and neutrino modes. **a)** Final-state π+ production for neutrino and electron scattering modes on carbon-12. **b)** Final-state π- production for neutrino and electron scattering modes on carbon-12.

4. Resonant Interaction Comparisons

Electron and neutrino scattering have been found to be quite similar for quasi-elastic events. However, as seen is Figure 4, DUNE is dominated by resonant events.

The process of resonant pion production is diagrammed in Figure 11. If we are to use electron scattering data to constrain reconstructions of DUNE neutrino events, it is necessary to understand the similarities and differences between neutrino and electron scattering modes in the resonant channel. In this way we can understand the extent to which the electron comparison is useful.

We compare various kinematic parameters of the scattering behavior of electrons and neutrinos in a GENIE simulation. Both sets of events were produced at 2.2 GeV with a carbon-12 target, and limited to the resonant channel. The kinematics of the two modes are quite similar. This lends further credence to the use of electron scattering data to constrain neutrino energy reconstruction models. However, there is a difference in the pions produced (Figure 12) largely due to the differences in charged current interactions between electrons and neutrinos.

5. Conclusions

The methods we have used here are effective for comparing reconstruction methods and working towards better results for DUNE. The calorimetric reconstruction seems to be quite accurate for neutrino events. We demonstrated that on a first pass, CLAS electron data holds potential to constrain neutrino reconstruction calculations. Simulated electron data for this purpose also seems to be promising. We have established a similarity between electron and neutrino scattering in the resonant channel, which can hopefully be used to extend the study with CLAS data to more types of DUNE events. Moving forward, the simulated DUNE data and feed-down calculations can be refined for more thorough reconstruction studies on more types of events. If refined, the use of electron data for neutrino reconstruction promises to vastly improve our ability to extract meaningful results from oscillation experiments and thereby deepen our understanding of neutrino interactions.

6. Acknowledgements

7. References

[1] Acciarri, Roberto et al. (2016). Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report Volume 1: The LBNF and DUNE Projects. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1601.05471.pdf

[2] Andreopoulos, Costas et al. (2015). The GENIE Neutrino Monte Carlo Generator: Physics and User Manual. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1510.05494.pdf

[3] Ashkenazy, Adi et al. (2018). Electrons for Neutrinos: Addressing Critical Neutrino-Nucleus Issues. Proposal to Jefferson Lab PAC 45. Retrieved on 24 November from https://www.jlab.org/exp_prog/proposals/17/PR12-17-006.pdf

[4] Esteban, Ivan et al. (2018). Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ23, δCP, and the mass ordering. Journal of High Energy Physics, 01(2019), pp. 106-141. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1811.05487.pdf

[5] Fantini, Guido et al. (2018). The formalism of neutrino oscillations: an introduction. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1802.05781.pdf

[6] Gil-Botella, Inés. (2015). Neutrino Physics. CERN Yellow Report, CERN-2013-003, pp. 157-206. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1504.03551.pdf

[7] Saraswat, K et al. (2017). Charged current quasi elastic scattering of muon neutrino with nuclei. Retrieved on 24 November 2019 from https://arxiv.org/pdf/1606.06916.pdf

Methods for Open and Reproducible Materials Science

Sara Laura Wilson[1], Micah Altman[2], Rafael Jaramillo[3]

1 Student Contributor, MIT Department of Materials Science and Engineering, Cambridge, MA 02139

2 MIT Center for Research on Equitable and Open Scholarship, Cambridge, MA 02139

3 Principal Investigator, MIT Department of Materials Science and Engineering, Cambridge, MA 02139

DATA STEWARDSHIP IN EXPERIMENTAL MATERIALS SCIENCE IS INCREASINGLY COMPLEX AND IMPORTANT. PROGRESS IN DATA SCIENCE AND INVERSE-DESIGN OF MATERIALS GIVE REASON FOR OPTIMISM THAT ADVANCES CAN BE MADE IF APPROPRIATE DATA RESOURCES ARE MADE AVAILABLE. DATA STEWARDSHIP ALSO PLAYS A CRITICAL ROLE IN MAINTAINING BROAD SUPPORT FOR RESEARCH IN THE FACE OF WELL-PUBLICIZED REPLICATION FAILURES (IN DIFFERENT FIELDS) AND FREQUENTLY CHANGING ATTITUDES, NORMS, AND SPONSOR REQUIREMENTS FOR OPEN SCIENCE. THE PRESENT-DAY DATA MANAGEMENT PRACTICES AND ATTITUDES IN MATERIALS SCIENCE ARE NOT WELL UNDERSTOOD. IN THIS ARTICLE, WE COLLECT INFORMATION ON THE PRACTICES OF A SELECTION OF MATERIALS SCIENTISTS AT TWO LEADING UNIVERSITIES, USING A SEMI-STRUCTURED INTERVIEW INSTRUMENT. AN ANALYSIS OF THESE INTERVIEWS REVEALS THAT ALTHOUGH DATA MANAGEMENT IS UNIVERSALLY SEEN AS IMPORTANT, DATA MANAGEMENT PRACTICES VARY WIDELY. BASED ON THIS ANALYSIS, WE CONJECTURE THAT BROAD ADOPTION OF BASIC FILE-LEVEL DATA SHARING AT THE TIME OF MANUSCRIPT SUBMISSION WOULD BENEFIT THE FIELD WITHOUT IMPOSING SUBSTANTIAL BURDENS ON RESEARCHERS. MORE COMPREHENSIVE SOLUTIONS FOR LIFECYCLE OPEN RESEARCH IN MATERIALS SCIENCE WILL HAVE TO OVERCOME SUBSTANTIAL DIFFERENCES IN ATTITUDES AND PRACTICES.

1. Introduction

At the intersection of data management, libraries, publishing, and machine learning lies the opportunity to develop the future of open and reproducible materials science. Combining these resources presents the potential to improve the publication, discovery, and use of experimental data.

Experimental materials science typically does not generate large quantities of data compared to fields such as geology, genomics, and economics. Ongoing, rapid progress in data science and the ever-increasing number of demonstrated applications of data science approaches in data-rich fields produce optimism that data science can be productively applied to materials science as well ("Technology: Sharing data in materials science," 2013). Significant progress in this direction requires significant data resources. Pioneering studies highlight the difficulty in assembling large quantities of experimental materials science data that can then be the basis for useful and insightful inferences (Kim, Huang, Jegelka, et al., 2017; Kim, Huang, Saunders, et al., 2017; Raccuglia et al., 2016). Data resources can grow through open-science practices such as sharing data generated across the research lifecycle, but experimental materials science lacks the norms, standards, and tools to make this widespread.

Open-science practices can also bolster the reputation and esteem of materials science in the broader society. Frequent replication failures in diverse scientific fields have drawn widespread attention and political chastisement (Fidler et al., 2018). Researchers, publishers, and funders increasingly recognize the need to characterize, test and evaluate practices for scientific replication, transparency, and data management (*Open Science by Design*, 2018). While the field of materials science has not been visited by a replication crisis, there are reasons to believe that the study of replication, transparency, and data management will yield benefits. The body of published literature in the field describes only a small subset of existing experimental results (Hill et al., 2016); more systematic curation and sharing of data and methods has the potential to enable data mining for the extraction of high-value knowledge (Kalidindi et al., 2015).

As building an infrastructure for open materials science is as much a social challenge as it is a technical one, we draw on conversations with and observations of Massachusetts Institute of Technology and Imperial College London faculty members to provide insight on the current data management practices in lab. We used a semi-structured interview method and developed a survey instrument to characterize three facets of data management: attitudes about sharing data, data sharing practices, and data stewardship practices. For details, see the extended white paper (Wilson et al., 2019).

An analysis of the interviews discussed by these experienced researchers suggests that although the importance of data sharing is broadly recognized, researchers in materials science vary in their practices and even definitions of data sharing. Further, the tools and practices highlighted in these discussions reveal the heterogeneity of data curation in this area, the relative dearth of scientific data management tools in use, and an absence of tools and practices for curating and sharing workflow.

2. Context: Data Sharing in Materials Science

Sharing data is increasingly recognized as critical to scientific practice and progress, but norms, practices, and attitudes vary widely across fields (Borgman, 2012). Information about data stewardship practices within labs can inform the design of an optimal system for data sharing between labs.

2.1. Current Data Sharing Practice

The published literature sharing of experimental data in materials science, beyond what is contained in publications, mainly takes the form of discrete uploads of parcels of data to a publicly accessible server. This is typically done at the time when a paper is accepted for publication and is done mainly for compliance with sponsor requirements. The actual work of sharing usually involves combing through experimental records, such as a large number of files in a compressed archive (e.g. a ZIP file), and copying the archive to a remote server. Data shared in this way often lacks context and is not easily discoverable.

Established platforms for sharing experimental materials science data were built as authoritative repositories. Well-known examples include the Landolt-Börnstein collection and the International Crystal Structure Database. These resources can provide trusted answers to questions like "What is the Vickers hardness of calcium titanate?", but they are not appropriate for warehousing unprocessed data and representing the scientific process. The value of sharing data from the complete research lifecycle has been recognized by the theoretical and computational materials science community. This community is developing platforms that can be used for designing computations, managing workflow, and storing the results such that all aspects of the work are continuously published. The Materials Project is the most well-known of these efforts, and there are many others (Hill et al., 2016; Jain et al., 2013). In contrast, the community of experimental materials scientists lacks the tools to meaningfully share data with colleagues and with the wider world. Recent efforts in this area include 4CeeD, which is designed to capture metadata originating from a centralized user facility, Materials Commons, which is designed to record materials processing steps for work on advanced metal alloys (Blaiszik et al., 2016; Nguyen et al., 2016; Puchala et al., 2016), and Materials Data Facility, which is designed as a general-purpose repository for materials data.

Sharing experimental data in materials science is made challenging by the nature of the research. Materials research data tend to be highly heterogeneous, small in volume (microscopy image data being an exception), and slow in acquisition rate (compared to other sciences and application realms of data science). Data is most often unlabeled, and meaningful interpretation requires knowing the processing history and other contextual information about the material in question. These qualities pose challenges both for sharing and for reusing experimental materials science data.

3. Methods

As our aim was to conduct an exploratory analysis of a small, selected group, we employed a semi-structured interview design (Creswell et al., 2011). Although this design requires more effort to implement than a close ended survey, it enables the discovery of practices and attitudes that are relevant to the questions asked but not specific to them. Further, it enables us to reflect on interviewee affect, especially concerning attitudes.

The goal in designing this survey is twofold. First, interviews will elucidate the various opinions on the benefits and drawbacks of data sharing and stewardship practices. By synthesizing this data, points of agreement and contention can be addressed. Second, discussion of data stewardship can help shape recommendations for future databases.

Data management practice in materials science were studied through in-person interviews with principal investigators (PIs) at two universities: Professor Kong, Dr. Ferralis, and Dr. Peters at the Massachusetts Institute of Technology (MIT), and Professor Walsh, Dr. Cairns, and Dr. Cooper at Imperial College London (ICL). All are early- and mid-career scientists. Each works either in the materials science department at his or her respective university or conducts research that directly contributes to and draws from data in materials science. They were chosen to represent the attitudes of the next generation of leaders in their fields, attitudes that may be expected to hold sway for the next several decades. Data sharing and stewardship practices within each lab, home university, and country inform their suggestions for an improved data management system. Diverse insight can be drawn from the selected faculty, as they participate in varied research within materials science. For example, in specific fields, established national and international standards govern data sharing practices while in others, data sharing varies between labs and research institutions.

Each faculty member was asked their perspective on the following topics: attitudes about sharing data, data sharing practices, and data stewardship practices. In conducting interviews, although subjects used their own terminology, we probed for responses using the following underlying definitions of these concepts broadly based on (Chervenak et al., 2000).

- *Data* encompasses any information pertaining to the work of science. This includes information traditionally entered in a lab notebook, files generated by computers running the instrument, instrument parameters, results of data analysis, and software used to perform the analysis. Data also includes the process (i.e. the workflow) of research, instead of just the results; this is especially important in materials science that is centrally concerned with processing materials. Data is generated at each step in the lifecycle of a research project, from initial inconclusive experiments to the final analysis presented in publication.

- *(Internal) Data stewardship* refers to the processes and standards used within the research group for controlled data creation, revision, management, integrity, and within-group collaboration.

- *(External) Data sharing* is defined as making data available to people outside of the collaboration that generated the data. Sharing can take the form of traditional publications, databases, websites, and personal communication. Thorough sharing requires that data are made discoverable, complete, accessible, and persistent: at this level, sharing starts to resemble publishing. Discoverability is the ability of a third party to learn of the existence of a shared data resource without explicit intervention by the first party (i.e. the researcher) or the second party (i.e. the organizing hosting the data server). Completeness is the extent to which shared data represents the full scientific cycle of learning, and usefulness is the extent to which the shared data can be used by a third party. Accessibility describes the ability of a third party to access the data once its existence has been discovered. Persistence describes the

length of time that shared data will remain discoverable and accessible. The extent to which data is discoverable, complete, accessible, and persistent informs the usefulness of data to third parties, and the attitude and reservations that researchers hold about sharing.

3. Interview Results

4.1. Professor Jing Kong

Interview Subject. Professor Kong, in the Department of Electrical Engineering and Computer Science at MIT, works on nanomaterials for electronics.

Data stewardship practices. In Professor Kong's lab, students and postdocs maintain their data individually.

Attitudes toward data sharing. Professor Kong notes that developing an understanding of collected data is as important as the data itself. She defines data to be everything collected during the exploration process, including but not limited to experiments, collaboration, and discussions. Data is the truth, even if it is counterintuitive to a preformed assumption. In a publication, the clear narrative develops an understanding of the data's significance in corroborating or debunking that assumption. Therefore, she claims that research papers are the pinnacle of data sharing because they provide meaning to the data through qualitative relationships.

Consequently, data in preliminary stages of research should not be shared. At this stage, the researcher may have an incorrect understanding of the data's significance, and she believes it would be counterproductive to share data with the wrong understanding. In this vein, she notes that the narrative or research is more important than the details, so it is unnecessary to share information in lab notebooks. It is impossible to read through all the papers that already exist, so sharing additional data would not be helpful.

A new system of data sharing would be helpful, Professor Kong mentions, if it provided a means by which failures could be shared and understood. Among publications, there is the incorrect perception that data must be successful, as in it supports a hypothesis or accomplishes a goal, to be published. She believes that developing an understanding of failures would be useful for future discovery.

4.2. Dr. Nicola Ferralis

Interview Subject. Dr. Ferralis is a research scientist the Department of Materials Science and Engineering at MIT. He is a member of the Concrete Sustainability Hub, which leads pioneering research on environmental-conscious concrete production.

Data stewardship practices. In Dr. Ferralis's lab, data maintenance is dependent upon the source of acquisition. If data is collected externally, such as at another university, the data is maintained within that university's system according to their policies.

For data collected internally, the lab maintains its own private server with functionality similar to that of Dropbox. The private server is required by companies from which the lab receives grants. This server maintains the lab's digital data, which is approximately 60-70% of acquired data, and can be accessed and downloaded by any lab member. The remainder of the data, such as experimental procedures, is stored individually in lab notebooks and personal records. To standardize data collection for the remaining 30-40%, the lab is constructing an infrastructure for record keeping. This infrastructure stores written experimental procedures and collected data, serving as a virtual lab notebook with the goal of understanding the function of each component of a process to more easily test and tweak multiple parameters.

Attitudes toward data sharing. Dr. Ferralis categorizes data as either acquisition from research, insight from personal experience, or insight from collaboration. An ideal research database, in his opinion, would include every component of data, as a small percent of researchers would find supplemental information to publications extremely useful. The challenge in providing everything is in developing a system with the ability to extract data in a way in which it is understandable to everyone. Since researchers have different habits for data collection, it must be either provided by the researcher in a standardized form or processed subsequently.

Dr. Ferralis notes two benefits of implementing a universal data maintenance system. First, a digital notebook would supplant the need for reports and would be a means to regularly track personal progress both internally and externally. Second, individual groups would not need to front the cost of managing a personal data storage site. The burden of a new system would be the initial resistance upon implementation. Courses and training on data management, like done in corporate research, would be necessary to advocate for the system. Once a critical mass of researchers have transitioned to the new system, a new culture of data science based upon standardized data maintenance can be achieved.

4.3. Dr. Ian Marius Peters

Interview Subject. Dr. Peters is a research scientist in the Department of Mechanical Engineering at MIT, and as of September 2019 is a group leader at Helmholtz Institute Erlangen, Nuremberg, Germany. At MIT he leads the "Accelerated Systems Development" team within the Photovoltaic Research Laboratory, which develops novel characterization and simulation methods for solar cells.

Data stewardship practices. In Dr. Peters's lab, data is maintained via multiple systems. Individuals save data to private computers and upload it to Dropbox to both backup and share it. The lab also has personal servers and hard drives, noting the security concerns of solely using online sites.

Attitudes toward data sharing. Dr. Peters claims that the most significant challenge in sharing data is distilling useful insights from the provided information. Papers, therefore, are the ideal conduit, as they provide a peer-reviewed, interpersonal, and logical product. Currently, if additional information is sought from a paper, the common strategy is to contact the author. While this step cannot be replaced, in Dr. Peters opinion, researchers would benefit from a complementary system for data sharing, in which information is provided to supplement papers. The platform would be a medium where researchers could share a concluding piece of work with the community, including both data sets and accompanying notes, in a format that can be referenced.

The concern in creating such a system is managing the data collection and selecting for which types of data should be shared. Researchers would benefit from a more structured approach to sharing code, rather than providing multiple channels, such as shared archives and personal websites. However, collecting everything, like a lab notebook, would be useless because it would be indecipherable without accompanying notes from the researcher. Providing this documentation would double or triple a researcher's workload. The challenge is to select for and handle data such that its usefulness outweighs the time penalty of documentation.

The ideal system would provide an interface which selects data sets that are useful for the user's project. A system which can

filter through, find, and access the right data for the user's research would be the most beneficial. The system must, therefore, have a selective internal structure that sorts, gathers, and formats data for optimal usefulness. The transition to this system will be difficult in universities, Dr. Peters notes. If it is mandatory, it must be enforced, and the resources to maintain it must be provided for by the university.

4.4. Professor Aron Walsh

Interview Subject. Professor Walsh works on computational materials science in the Department of Materials at Imperial College London. He is head of the Materials Design Group, which develops theoretical solutions for applied problems in solid-state chemistry.

Data stewardship practices. In Professor Walsh's lab, data is open and accessible to all lab members. For small projects, the lab utilizes Box for data storage, which Imperial has an institutional agreement with for unlimited storage. For larger projects, the group maintains individual servers. Professor Walsh's group benefits from open data sharing by performing analysis on large databases; they look for correlation in others' through post-processing without needing to reproduce the data.

Attitudes toward data sharing. Professor Walsh states that in Europe, there is a strong push towards open data and science. EU and UK funders have data access statements, and journals require raw data associated with the paper to be made publicly available. Data sharing is heterogeneous depending on the file type and size. For example, Zenodo is a data repository in Europe for the input and output files of a code necessary to reproduce the computational experiment. For custom tools, such as post-processing code designed to create the figures in the paper, the code is made available on code sharing platforms like Github. However, challenges exist among existing data repositories. There is no community census for data sharing. Zenodo is difficult to search because there are mixed file formats. Since the data is not curated, it is difficult to distinguish between accurate and incorrect numbers. Increased maintenance would improve existing databases by determining a data's validity through comparison to existing data. An ideal system for data sharing would curate all information associated with a publication and make it publicly available.

Initiatives to standardized data sharing are underway, such as the Nomad Project, which collects data from research groups and converts it to a common format that is searchable and available to everyone. The field of crystallography has the most sophisticated databases in materials science. A common file format, CIF (Crystallographic Information File), was established in the 1990s that is both human and machine readable. Crystallography databases can check the file against standard content to determine if it is complete and accessible. Data that is falsified or structures that are unphysical is easily sorted.

To develop a standard for labs to adopt, a large critical mass of researchers need to agree upon a specific file format. Professor Walsh suggests agreement among the leaders of a field will create a shift among all groups in that field. The transition will be difficult because groups will be reluctant to back-track to arrange data in a correct format. Publishers and funders need to push to make data sharing mandatory. However, once a standard format is established, data will be more accessible for replication and post-processing by external groups.

4.5. Dr. Andrew Cairns

Interview subject. Dr. Cairns works on high-pressure crystallography in the Department of Materials at Imperial College London. His lab extrapolates data from crystallographic imagery to develop materials with unique properties.

Data stewardship practices. In his lab, data is stored on Box. He notes that it is easiest to use the system provided by the university because it is standard among all the research groups. Everyone within the lab has access to all data produced, and each research is responsible to maintain his or her data with personalized organization and labeling. This maintenance is not standardized among the members.

Attitudes toward data sharing. Dr. Cairns observes that there is no consistency among external data sharing systems in materials science. Large databases exist, like the Cambridge Structural Database, which is searchable and open for further analysis upon the data. However, post-processing is only possible if every intermediate step leading to the final result is accessible. Enough metadata needs to be published such that a researcher can both reproduce the result and extract new meaning from the data with different analysis tools.

Dr. Cairns is a crystallographer, and he notes that the success of crystallographic databases was the result of a community-wide effort and funding. In crystallography, the expectation of uploading to a database is cultural, so researchers prepare their data as it is collected. Checks related to statistical measures and methods for files exist in databases as a barrier to falsifying data. The long-term advantage of this is tangible because this organization has allowed databases to be searchable, which is much easier than reading the prose of a research paper and transcribe numbers oneself. Progress is underway in other fields to establish similar systems. In the field of semiconductors and dielectrics, companies are paying researchers to read through old papers and make databases with a consistent format so the data can be used for machine learning.

The Cambridge Crystallographic Data Center (CCDC) and Inorganic Crystal Structure Database (ICSD) are the two main databases used by crystallographers. The databases are useful because rules exist such that the researcher must provide every piece of information in a standard format. The database has a searchable interface and researchers can then easily extract data from this format. This CCDC has data management training programs, which can be purchased. Data uploading to the CCDC occurs at the time of publication, in which a corresponding reference number is provided with the papers. Other policies, such as that of the European Synchrotron Radiation Facility (ESRF), require the release data for open use after three years, regardless of whether a publication has been released.

4.6. Dr. Sam Cooper

Background. Dr. Sam Cooper in the Dyson School of Design Engineering at Imperial College London uses both machine learning and experimental techniques to research electrochemical energy storage devices such as batteries and fuel cells. Dr. Cooper uses both machine learning and experimental techniques to research electrochemical energy storage devices such as batteries and fuel cells.

Data stewardship practices. For experimental techniques, like 3D imaging, tens of gigabytes of data are produced per each scan and processing step. Centralized college storage services offer solutions for quantities of data an order of magnitude below that

necessary for machine learning, so in Dr. Cooper's lab, data is stored on two image analysis machines with RAID 2 (Redundant Array of Independent Disks) memory blocks. Data sharing within the lab is entirely open. Dr. Cooper acknowledges that the associated overhead in creating a data management system specific to the lab is an expensive invest in both money and time, and it is therefore often required of the individual to maintain the cleanliness of his or her data.

Dr. Cooper notes that sharing data with a publication significantly improves the researcher's ability to have a highly cited and impactful paper; his two most influential works were pieces of open source software. Researchers often look at the supplementary material in a publication for data, but this data is only useful if it is complete. Dr. Cooper defines complete as enough information to replicate the experiment.

Attitudes toward data sharing. Dr. Cooper asserts that rather than develop a new database for researchers, publications themselves could be restructured. Journals offer a researcher prestige, but publications are not useful for extracting data. Dr. Cooper suggests two forms of papers should replace the current style of publication. First, systematic review papers, in which specific search terms are identified across the referenced publications, would be as useful as a searchable database. Through comparison in these review papers, errors in specific data sets could be distinguished. Second, data should not be forced into the mold of a publication. The structure of a paper should be reformatted to minimize the background section and emphasize the new data. Information that is useful for data sharing and replication includes measurement techniques and the raw data connected with figures. Journals should be responsible for requiring this data, teaching the new system, and provided standards for published data. Resistance may occur, as requiring more data will slow the rate of data publication, which is counterintuitive for for-profit publication models.

An addition to these new form of publication papers would be a paper-commenting system, like Reddit for publications. Dr. Cooper notes that papers are the most inefficient form of communication because each reader must read it as if he or she is the first reader, not in the light of everyone who has read it before and found mistakes. Within the machine learning community, there is no systematic approach to managing and sharing data. Data is shared heterogeneously across multiple platforms. The ideal database would be a Google Scholar equivalent that is a search engine for scientific values with an iterative approach involving key terms and variables provided by the searcher.

5. Discussion and Recommendations
5.1. Discussion of survey results

All respondents showed a deep well of feeling for the topic of data sharing and open science, and a common frustration that existing systems are inadequate (or, as infomercials put it, "There has to be a better way!"). However, respondents diverged substantially in the values they place on data sharing and open science in their current practices and hopes for the future. This diversity of opinion is a caution against one-size-fits-all approaches.

Data sharing presents burdens and opportunities. Two contrasting opinions exist: data sharing activities are thought to provide no value to the research process, either over the short or long term, and are done for compliance with norms or explicit sponsor requirements. Conversely, by sharing enough data to allow your experiments to be reproduced, it will broaden one's

impact and standing in the field. Attitudes on sharing data inform the amount of research scientists will be willing to share. All interviewees agreed that data sharing is in dire need of proper management and that data should be shared. Interviewees agreed further that sharing data is useful only if that data is discoverable and universally readable.

There is substantial diversity with respect to the point during research at which data should be shared, the quantity of data shared, and even the definition of "data". Some argue that research should be shared at the point of publication, justifying this statement by claiming that the publication provides a complete, correct narrative that contextualizes the data. Additionally, openly providing research data prior to publication jeopardizes the publication's impact. The counterargument, noted by some researchers, is that data sharing at the point of acquisition provides insight for the research, which functions as a form of collaboration to further the research.

The most notable dissent was expressed in desirable amount of data to shared. Some respondents opined that providing all data (broadly defined) should be shared; it would be useful for the small portion of the population in need of detail to replicate experiments, and extra data would, in turn, increase the impact of the publication. Other respondents argue that no data sharing is necessary beyond the publication, as quantity of publications in existence already exceeds that which is capable of being read. Notably, in these respondents simultaneously asserted that data sharing was vital: this implies that they included publications within their definition of "data," which deviates from the common use of the term.

Data stewardship practice among the respondents was informal, heterogenous, and coarse-level. Outside of crystallography, respondents made no mention of data citation, identifiers, metadata standards, or workflow curation. The dominant tools mentioned provide basic backup and internal dissemination only and only at the level of aggregation of file and folder. Common curation affordances and integration with reproducible science tools and workflows were generally absent. The tools mentioned did not provide revision control, persistent identifier generation, and metadata creation nor did they integrate with tools for lab notebook management, workflow or instrument management, or report and publication. Overall, these stewardship practices appears to lag state-of-the-art in bench science (Lowndes et al., 2017).

5.2. Recommendations

The practice of sharing data externally only at the time of publication and when required by the journal is a de facto default, but it comes with risks. Delaying preparation for data sharing until requested reduces reproducibility (Wicherts et al., 2011). Further, sharing only data related to journal publication exacerbates file-drawer problems at disciplinary level; work is repeated, anomalies emerge more slowly, and effect size may be overestimated (Dwan et al., 2013).

Based on this analysis, we conjecture that the following would benefit the field without imposing substantial burdens on researchers. Data sharing should be comprised of two components. The first is the broad adoption of basic file-level data sharing at the time of manuscript submission. The submission should be a regulated format with specific details required, such as data collection and processing methods, to aid in reproducibility. The aim of presenting the data in a specific format is to increase its usability by presenting the data in a readable format to facilitate

use by a third-party. Requirements for formatting and metadata uniformity may lessen in time as automated software becomes more adept at interpreting diverse data sources (Kim, Huang, Saunders, et al., 2017; Satheesan et al., 2018; Tshitoyan et al., 2019).

The second component is software for managing data during research. This would be a highly developed laboratory information management system (LIMS) encompassing electronic notebooks, workflow tracking, raw and processed data file curation, supply management, and so on. Such an all-encompassing research data management system - an omniscient, digital research secretary - likely will remain a dream for some time, at least for academic research. However, key elements required for data sharing and open science can be implemented with much less effort. For instance, data file tracking software can be overlaid on of folder-level tools such as Box and Dropbox to facilitate intra-group sharing prior to publication. The data management system should require minimal documentation for saved information (i.e. metadata) that can enable data sharing at the time of article submission to becomes seamless, with minimal marginal effort on the part of the researchers.

There are clear needs for improved data management tools, but transitions will remain difficult due to the diversity of hopes, expectations, and current practices in experimental materials science. The connections between research data and publications could be strengthened and improved by a common set of expectations for data sharing, more nuanced than current requirements but sufficiently general that different researchers could use different software tools to comply. Such expectations could be developed by a convening of publishers, materials scientists and their member organizations, research librarians and administrators, and other stakeholders. A useful step in this direction would be greater emphasis on teaching data management at the graduate and undergraduate level, both to directly communicate best-practices and to create more participatory forums in which to debate and define the future of data sharing and open science.

6. Author Contributions

The authors describe contributions to this article using a standard taxonomy (Allen et al., 2014). M.A. and R.J. provided the core formulation of the research goals and aims and methodology. R.J. lead administration, supervision. S.W. led the data collection, investigation, and the writing of the original manuscript. All authors contributed through commentary, review, editing, and revision.

7. Acknowledgements

8. References

[1] Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. Nature, 508(7496), 312–313. Retrived on August 13, 2019 from https://doi.org/10.1038/508312a

[2] Blaiszik, B., Chard, K., Pruyne, J., Ananthakrishnan, R., Tuecke, S., & Foster, I. (2016). The Materials Data Facility: Data Services to Advance Materials Science Research. JOM, 68(8), 2045–2052. Retrived on August 13, 2019 from https://doi.org/10.1007/s11837-016-2001-3

[3] Borgman, C. L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6), 1059–1078. Retrieved on August 13, 2019 from https://doi.org/10.1002/asi.22634

[4] Chervenak, A., Foster, I., Kesselman, C., Salisbury, C., & Tuecke, S. (2000). The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. Journal of Network and Computer Applications, 23(3), 187–200. Retrieved on August 13, 2019 from https://doi.org/10.1006/JNCA.2000.0110

[5] Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research. SAGE Publications. Retrieved from Retrieved on August 13, 2019 from https://in.sagepub.com/en-in/sas/designing-and-conducting-mixed-methods-research/book233508

[6] Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., & Reporting Bias Group. (2013). Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. PLoS ONE, 8(7), e66844. Retrieved on August 13, 2019 from https://doi.org/10.1371/journal.pone.0066844

[7] Fidler, Fiona, & Wilcox, John (2018). Reproducibility of Scientific Results. The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. Retrived on August 13, 2019 from https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=scientific-reproducibility&archive=win2018.

[8] Hill, J., Mulholland, G., Persson, K., Seshadri, R., Wolverton, C., & Meredig, B. (2016). Materials science with large-scale data and informatics: Unlocking new opportunities. MRS Bulletin, 41(5), 399–409. Retrieved on August 13, 2019 from https://doi.org/10.1557/mrs.2016.93

[9] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., … Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002. Retrieved on August 13, 2019 from https://doi.org/10.1063/1.4812323

[10] Kalidindi, S. R., & De Graef, M. (2015). Materials Data Science: Current Status and Future Outlook. Annual Review of Materials Research, 45(1), 171–193. Retrieved on August 13, 2019 from https://doi.org/10.1146/annurev-matsci-070214-020844

[11] Kim, E., Huang, K., Jegelka, S., & Olivetti, E. (2017). Virtual screening of inorganic materials synthesis parameters with deep learning. Npj Computational Materials, 3(1), 53. Retrieved on August 13, 2019 from https://doi.org/10.1038/s41524-017-0055-6

[12] Kim, E., Huang, K., Saunders, A., McCallum, A., Ceder, G., & Olivetti, E. (2017). Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. Chemistry of Materials, 29(21), 9436–9444. Retrieved on August 13, 2019 from https://doi.org/10.1021/acs.chemmater.7b03500

[13] Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., … Halpern, B. S. (2017). Our path to better science in less time using open data science tools. Nature Ecology & Evolution, 1(6), 0160. Retrieved on August 13, 2019 from https://doi.org/10.1038/s41559-017-0160

[14] Nguyen, P., Konstanty, S., Nicholson, T., O'brien, T., Schwartz-Duval, A., Spila, T., … Paquin, N. (2016). 4CeeD: Real-Time Data Acquisition and Analysis Framework for Material-related Cyber-Physical Environments. Retrieved on August 13, 2019 from https://www.ideals.illinois.edu/handle/2142/91007

[15] Open Science by Design: Realizing a vision for 21st century research. (2018). Washington, D.C.: National Academies Press. Retrieved on August 13, 2019 from https://doi.org/10.17226/25116

[16] Puchala, B., Tarcea, G., Marquis, E. A., Hedstrom, M., Jagadish, H. V., & Allison, J. E. (2016). The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community. JOM, 68(8), 2035–2044. Retrieved on August 13, 2019 from https://doi.org/10.1007/s11837-016-1998-7

[17] Raccuglia, P., Elbert, K. C., Adler, P. D. F., Falk, C., Wenny, M. B., Mollo, A., … Norquist, A. J. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature, 533(7601), 73–76. Retrieved on August 13, 2019 from https://doi.org/10.1038/nature17439

[18] Satheesan, S. P., Marciano, R., Marini, L., Minsker, B. S., Navarro, C. M., Schmidt, A., … Lee, J. (2018). Brown Dog: Making the Digital World a Better Place, a Few Files at a Time. In Proceedings of the Practice and Experience on Advanced Research Computing - PEARC '18 (pp. 1–8). New York, New York, USA: ACM Press. Retrieved on August 13, 2019 from https://doi.org/10.1145/3219104.3219132

[19] Technology: Sharing data in materials science. (2013). Nature, 503(7477), 463–464. Retrieved on August 13, 2019 from https://doi.org/10.1038/503463a

[20] Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., … Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature, 571(7763), 95–98. Retrieved on August 13, 2019 from https://doi.org/10.1038/s41586-019-1335-8

[21] Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE, 6(11), e26828. Retrieved on August 13, 2019 from https://doi.org/10.1371/journal.pone.0026828

[22] Wilson, S. L., Altman, M. & Jaramillo, R. (2019). Methods for Open and Reproducible Materials Science, SocArXiv. Retrieved on November 21, 2019 from https://doi:10.31235/OSF.IO/AG8ZU

Iman Jilani

Adekunle Onadipe
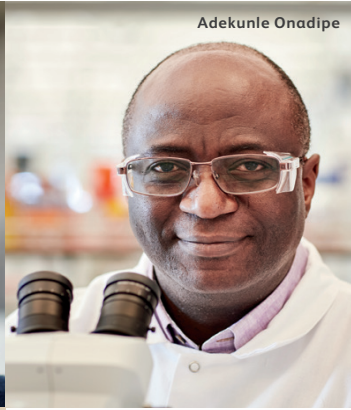
Guoyun Bai

Ricky Fernandes

Herbert Medina

Sonal Bhatia

Everything that makes us unique makes us uniquely good at the work we do together.

WE ARE THE MANY DARING, DIFFERENT PEOPLE OF PFIZER

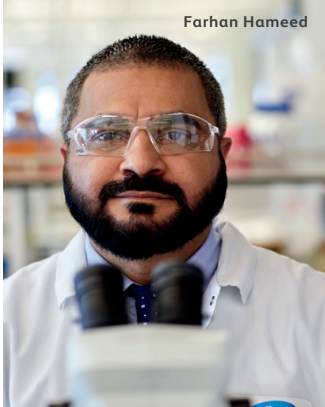## ALL WORKING TO DELIVER BREAKTHROUGHS THAT CHANGE PATIENTS' LIVES.

Regina McDonald

Charles Cain

Pfizer

Farhan Hameed

Karen Walters

Come and discover what we are all about at

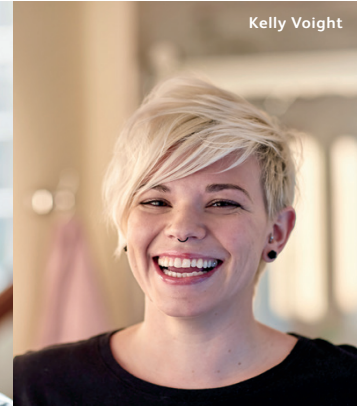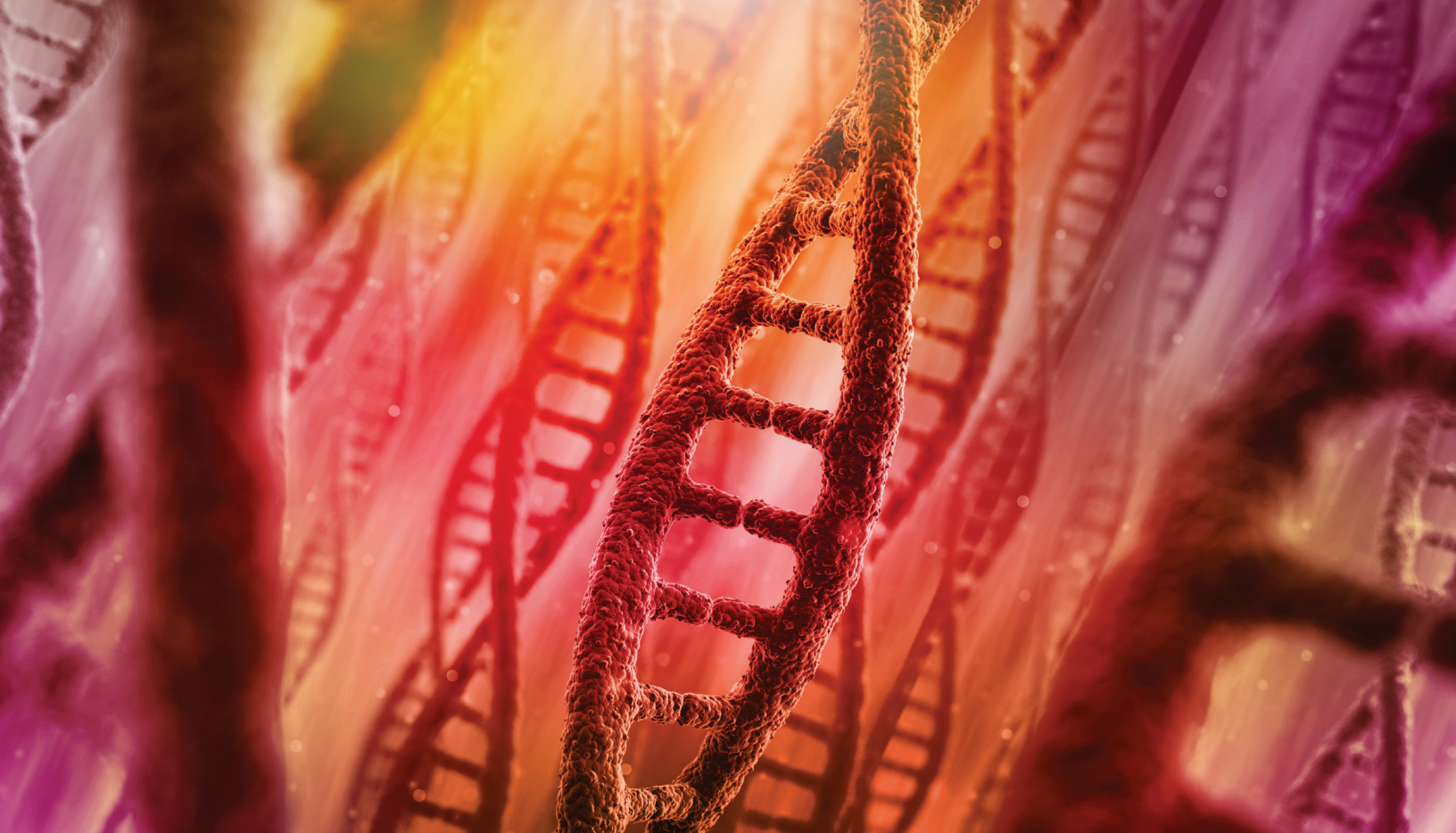**pfizer.com/careers**

Elaine Ravasco

Adekola Alagbe

Kelly Voight

Patrick McCann

Better Health, Brighter Future

At Takeda Oncology, we endeavor to deliver novel medicines to patients with cancer worldwide through our commitment to science, breakthrough innovation and passion for improving the lives of patients.

This singular focus drives our aspirations to discover, develop and deliver breakthrough oncology therapies. By concentrating the power of leading scientific minds and the vast resources of a global pharmaceutical company, we are finding innovative ways to improve the treatment of cancer.

We've built a portfolio of paradigm-changing therapies and a leading oncology pipeline. Though we've made great strides in our fight against cancer, we are determined to do more – to work harder and to reach higher. We continue to seek our aspirations with the same passion, agility and entrepreneurial spirit that has sustained our patient-centric culture and has made us the leaders in oncology that we are today.

We know that our mission is not a quick or simple one, but we are up for the task: **we aspire to cure cancer**.

🐦 📘 📷 👻 in ▶ 🎵 @TakedaOncology

*To learn more, visit us at **takedaoncology.com.***

Takeda | ONCOLOGY