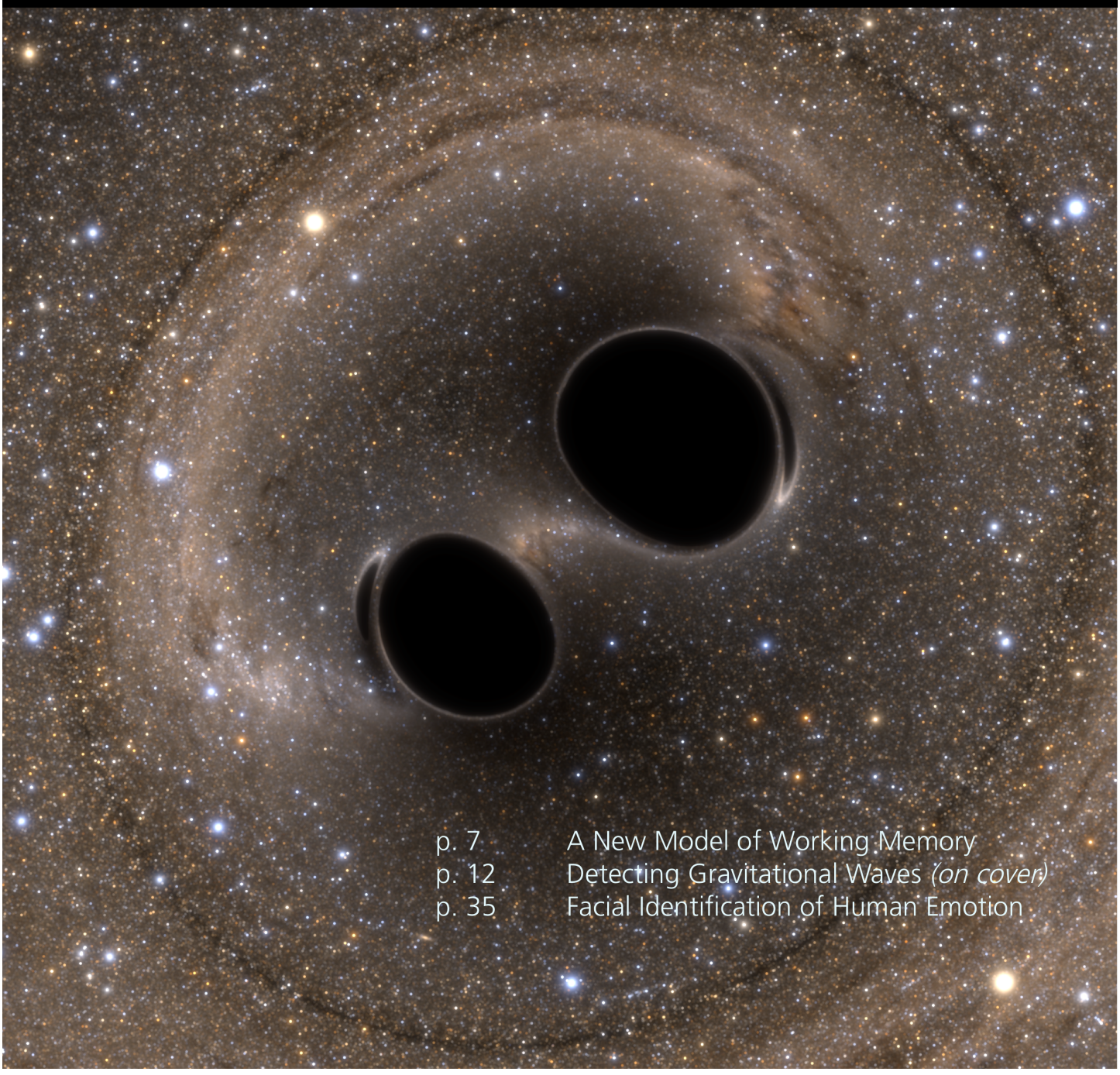
Volume 31 Spring 2016

MURJ

Massachusetts Institute of Technology
Undergraduate Research Journal



naturejobs.com

Eager to move on up in your career?

Naturejobs is the global jobs board and career resource for scientists. We can help you throughout your job search and your career:

Find a job
Search jobs, set up job alerts and research employers or search for jobs on-the-go with our free mobile app.

Help employers to find you
Upload your CV and make your profile searchable to employers.

Meet employers in person
Attend the *Naturejobs* Career Expo for invaluable career advice and to meet recruiters.

View science careers advice
Keep up with the latest careers articles, interviews and more via our news and resources section or by subscribing to our newsletter.

Ask us questions
Search for "*Naturejobs*" on your preferred social media platform or contact us via the *Naturejobs* blog.

naturejobs.com

Follow us on:

SPRINGER NATURE

MURJ

Massachusetts Institute of Technology Undergraduate Research Journal

Contents

Massachusetts Institute of Technology

UNDERGRADUATE RESEARCH JOURNAL
Volume 31, Spring 2016

MURJ Staff
MIT Undergraduate Research Journal

Massachusetts Institute of Technology

May 2016

Dear MIT community,

We are excited to publish the 31st issue of the MIT Undergraduate Research Journal, a biannual student-run publication that features groundbreaking undergraduate research from across campus and across disciplines. As always, the research summaries and articles presented in our journal reflect the extraordinary talent and passion of MIT undergraduates. In this issue, we learn about an autonomous agent communication model designed to facilitate effective human-robot team communication, an Automatic Emotion Detection System developed to identify human emotions in facial expressions, and a cutting-edge cryptographic method with wide-ranging immediate applications.

In addition to our research articles, we also present two features articles that explore groundbreaking scientific advances. In this issue, one features article explores the detection of gravitational waves, a revolutionary discovery by LIGO that hails the beginning of a new era in astronomy. In addition, we present an interview with Professor Hidde Ploegh, a member of the Whitehead Institute for Biomedical Research and a Professor of Biology at MIT.

Biannual publication of this journal is a collaborative undertaking by an extraordinary team of dedicated students. We would like to thank our editorial board and contributors for their time and effort this semester. In addition, we would like to thank all of the undergraduates who shared their research with us and the greater MIT community.

Massachusetts Institute of Technology

MURJ Staff
MIT Undergraduate Research Journal

If you are interested in contributing to future issues of the MIT Undergraduate Research Journal, we invite you to join our team of authors and editors or submit your research for our Fall 2016 issue. For previous issues of this journal, visit our website at murj.mit.edu. Please contact murj-officers@mit.edu should you have any questions or comments.

Best,

Tatyana Gubin
(Co-Editor-in-Chief)

Linda Jiang
(Co-Editor-in-Chief)

Lakshmi Subbaraj
(Co-Editor-in-Chief)

Massachusetts Institute of Technology

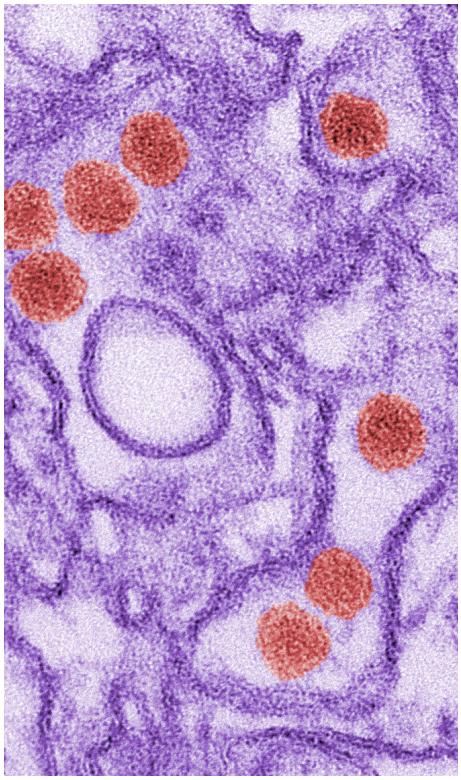
UNDERGRADUATE RESEARCH JOURNAL
Volume 30, Fall 2015

MIT Science News In Review

Memory | Zika Virus | Emotion | Cancer | Bacterial Growth

Public Health

MIT Tackles Zika Virus



Digitally-colored transmission electron micrograph (TEM) of the Zika virus (in red), which is 40 nm in diameter.

(Source: Wikimedia, Photo credit: Cynthia Goldsmith)

On February 1, the World Health Organization (WHO) declared the spread of the Zika virus disease an international public health emergency. Transmitted via mosquitoes, the Zika virus disease exhibits symptoms similar to its relative, Dengue fever, such as fever, joint pain and conjunctivitis. However, unlike Dengue fever virus, Zika virus has been found to be linked to microcephaly, a condition in which babies are born with small heads and brain damage. A link between Zika virus and Guillain-Barré syndrome, a condition in which the immune system attacks the nerves, is strongly suspected but not yet confirmed.

At the forefront of the battle against the Zika virus, Professor Lee Gerkhe of MIT's Institute for Medical Engineering and Science (IMES) and his lab are spearheading the efforts for improved disease diagnostics. In the past, the Gerkhe lab has developed successful non-cross-reacting diagnostic tests for the Dengue fever virus and for ebolavirus.

The group's goals are to develop sensitive, low cost, rapid diagnostics to detect Zika virus with immediate results and to conduct sufficient research to characterize Zika virus infections. Gerkhe notes that one challenge is that the Zika virus has the potential to spread across vast geographic distances via mosquito vectors. Furthermore, Zika disease can be asymptomatic, making it difficult to identify and contain the spread of disease. Gerkhe's lab hopes to provide to health workers with a test that would not only identify those who have Zika virus but also those who have carried the disease in the past, which could be done by developing a method to identify the presence of Zika-specific IgM and IgG antibodies in patients.

Gerkhe's lab is also focused on the development of low-cost communication and data technologies to acquire and transmit diagnostics data to physicians and health workers, as well as "smart" mosquito traps that transmit data on the species of captured mosquitoes and potentially even screen the mosquitoes for the presence of Zika. This would allow communities to effectively prepare for the onslaught of an epidemic.

Meanwhile, Daniel Anderson, an associate professor of chemical engineering and a member of the IMES, has been working to develop a customized, on-hand, single-dose RNA nanoparticle vaccine that would contain RNA to train immune systems to fight against certain instructed pathogens. Thus far, the Anderson lab has been successful in developing customized vaccines for various "priority pathogens" that act against animal immune systems.

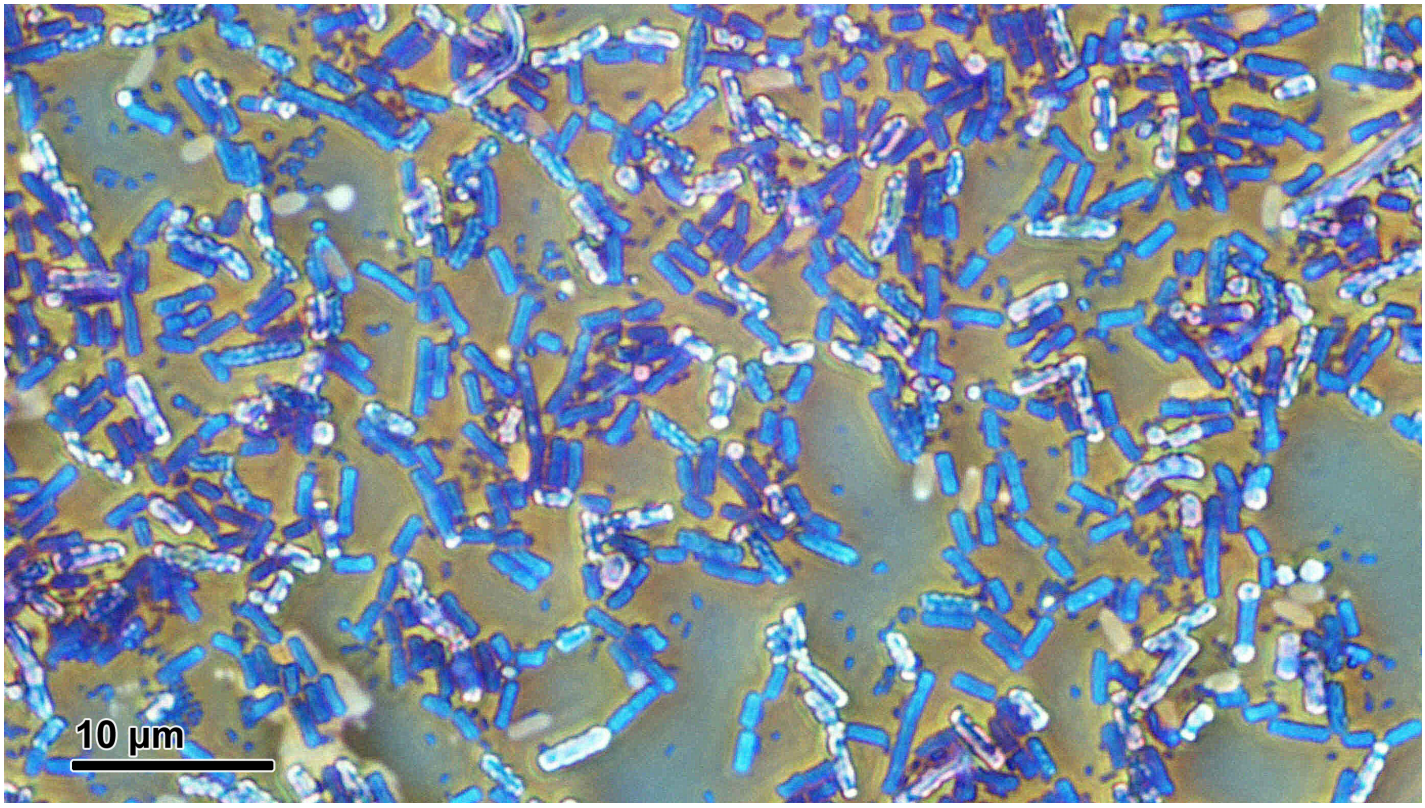
The Anderson lab is currently developing a similar vaccine mechanism for the Zika virus disease. This method is particularly compelling as the speed, cost and local production potential of the vaccine allows for effective treatment of the disease once Zika-specific antigens are identified by virologists. Furthermore, the nanoparticle has the capability to simultaneously target multiple diseases associated with mosquito vectors, such as ebolavirus.

According to Anderson, if this vaccine is successful during clinical trials, it will be the "first and only nonviral replicon delivery system that has achieved protective immunity in lethal exposure experiments" without any risk of the immune system attacking the vaccine.

—T. Alam

BIOPHYSICS

Bacterial Patch Formation



Stained spores of *Bacillius subtilis*. (Source: Wikimedia, Photo credit: Joseph Reischig.)

Nature is a keen study of the adage "safety in numbers"—a variety of microorganisms, including algae and stromatolites, exhibit a tendency to cluster together. The bacterium *Bacillus subtilis* is no exception to this trend: researchers in the Physics of Living Systems group found that *B. subtilis* populations promote survival by forming patches in resource-poor environments. *B. subtilis* consumes glucose. When it is unable to obtain glucose directly, it cooperatively digests more complex starches into glucose using the enzyme amylase.

Christoph Kratzke, a postdoctoral student in the Physics of Living Systems group, cultured the bacteria in well-mixed environments containing both simple and complex sugars in order to elucidate what causes the clustering behavior. Kratzke and Jeff Gore, the Latham Family Career Development Associate Professor of Physics, observed patch growth in the latter. To eliminate better food-sensing techniques or higher mobility as reasons for patch formation, the researchers also cultured mutants with impaired abilities. The bacteria without food-sensing mechanisms or mobility spontaneously formed patches in a medium with complex sugars.

The key advantage of *Bacillus subtilis*' clustering lies in the property of cell density. In a well-mixed environment, bacteria are not able to survive unless a certain threshold cell density is met. In an unmixed environment, *B. subtilis* was able to survive at much lower cell densities due to patch formation. Kratzke and Gore further found that survival and population growth interacted with each other in a compelling way: in environments where bacteria could easily move around and expand, they formed fewer patches and so could not cooperate as effectively. This in turn negatively affected their growth. *B. subtilis*' patch formation and balancing act between survival and expansion could have ecological implications, and may provide insight into similar behaviors in other species.

—N. SYED

Neuroscience

The Role of Neurons in Processing of Emotions

The myriad events in our daily lives trigger emotional responses, and a mismatch between an event and its corresponding response may underlie mental illness. In the interest of investigating this idea, researchers at the Picower Institute for Learning and Memory examined how two neuron populations play a role in emotional processing. Kay Tye, the Whitehead Career Development Assistant Professor of Brain and Cognitive Sciences, and her colleagues Anna Beyeler and Praneeth Namburi investigated neurons in the amygdala that send information to the nucleus accumbens, which corresponds to reward, and the centromedial amygdala.

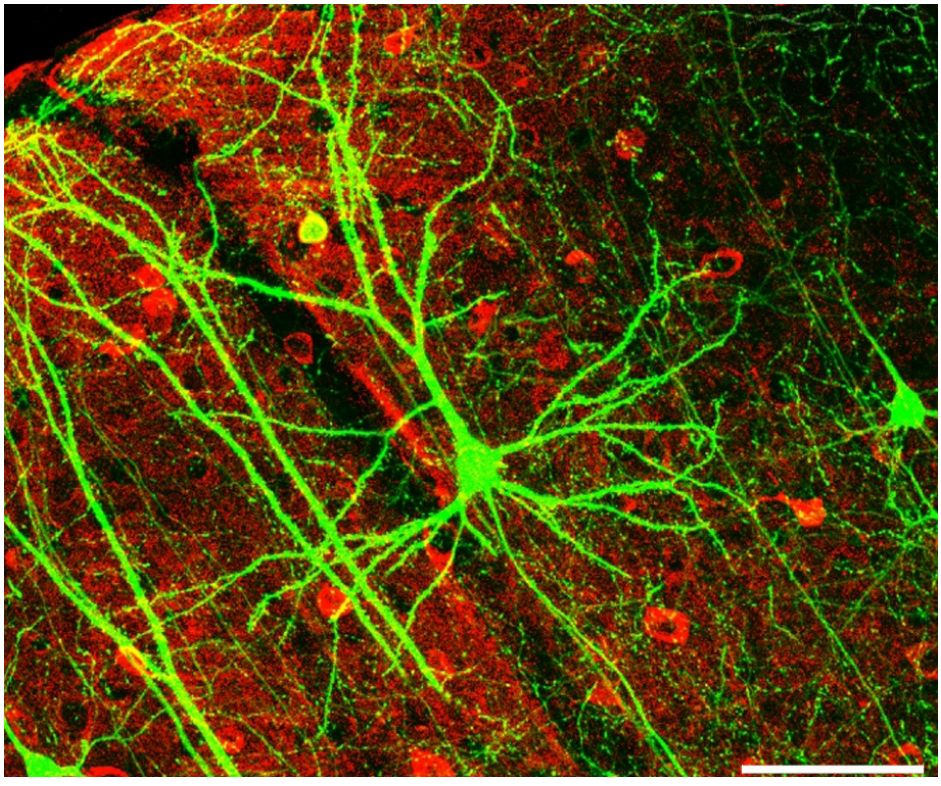
The researchers tagged these two neuron types, as well as neurons corresponding with the ventral hippocampus, in two groups of mice. Channelrhodopsin, a light-sensitive protein, provided Tye and her colleagues with the ability to distinguish between neuron populations. The mice were conditioned to associate two tones with a pleasant and bitter taste, respectively.

Researchers noted electrical activity in the neuron populations as the mice heard the two tones. Among the three cell populations, not all neurons reacted in the same manner, but Tye and her colleagues discerned three main trends. Neurons that relayed information to the nucleus accumbens largely responded to tone associated with a pleasant taste, while those associated with the centromedial amygdala reacted to the one associated with an unpleasant taste. The neuron population that sent information ventral hippocampus responded to both tones.

The findings regarding these three cell populations in the amygdala may mark a transition in how neuroscientists study emotions. Previously, the main focus was on brain regions, but Tye and her colleagues indicate that neuron populations are a new area of interest. The fact that such varied neurons reside in one brain region is also worthy of further investigation. Tye hypothesizes that this proximity may enable neurons to better communicate with each other and in turn drive our ability to respond quickly to stimuli.

—N. Syed

Pyramidal neurons, found in the hippocampus.
(Source: Wikimedia, Photo credit: Wei Chung Allen Lee, et. al.)



Oncology

Unraveling the Molecular Mechanisms of Chemotherapeutic Resistance in Cancer

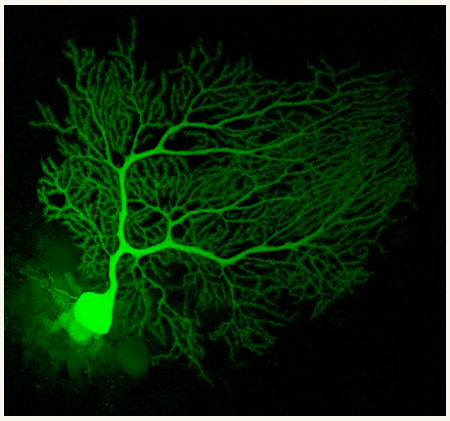
One of the biggest challenges in cancer treatment is the heterogeneity of a patient's response to chemotherapeutic drugs, and a great portion of cancer research is focused on understanding the molecular mechanisms that give rise to this heterogeneity. In a paper published in the journal Cancer Discovery, researchers at MIT and MGH investigate why kinase inhibitors, a class of drugs used to treat many types of cancers, fail to produce their intended results in many patients. In a previous study, researchers from the Lauffenberger lab had observed that when endometrial cells were treated with kinase inhibitors, backup systems that

NEUROSCIENCE

A New Model of Working Memory

Working memory, the system that enables us to hold information in our mind for immediate use, was long thought to function fluidly: sets of neurons associated with the necessary information would fire continuously, allowing for its storage and retrieval. A study conducted by researchers at the Picower Institute for Learning and Memory provides evidence to question such a model.

Animals observed sequences of colored squares, which were manipulated such that one square switched color. In order to successfully complete the task, the animals had to store information about each of the squares in working memory. Researchers Ed Miller, Mikael Lundqvist, and Jonas Rose recorded the animals' neuronal activity during the experiment. They noted that neurons in the prefrontal cortex fired intermittently, a phenomenon in direct contrast with previous observations of working memory. Prior models of working memory stemmed from averaging neuronal activity, and gave the impression that sets of neurons fired continuously.

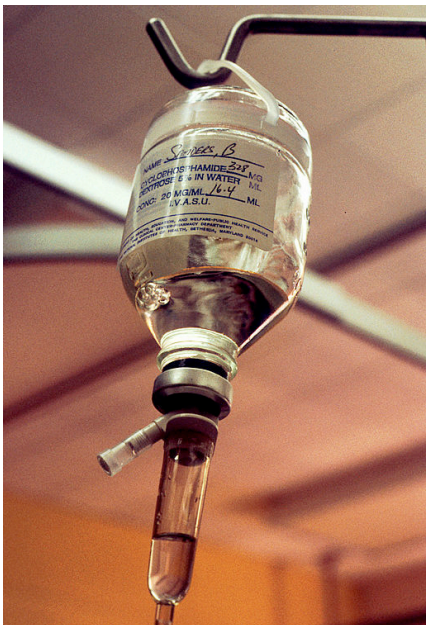


Research suggests that neurons contributing to the formation of working memory do not fire continuously.
(Source: Wikimedia, Photo credit: Maryann Martone)

The periodic behavior of neurons in working memory aligns with an alternative model that Lundqvist formed prior to the study. Lundqvist's model proposes that information is stored in bursts among groups of neurons that keep recurring for as long as that information is needed. These bursts combine to form gamma waves that correspond to a certain object in memory.

The part-and-parcel nature of neuronal activity in working memory could indicate that other processes within the brain, such as attention, follow a similar pattern. The apparent smoothness of our minds may in fact belie compartments of information flashing between neurons—a compelling new conception of our inner workings.

—N. SYED



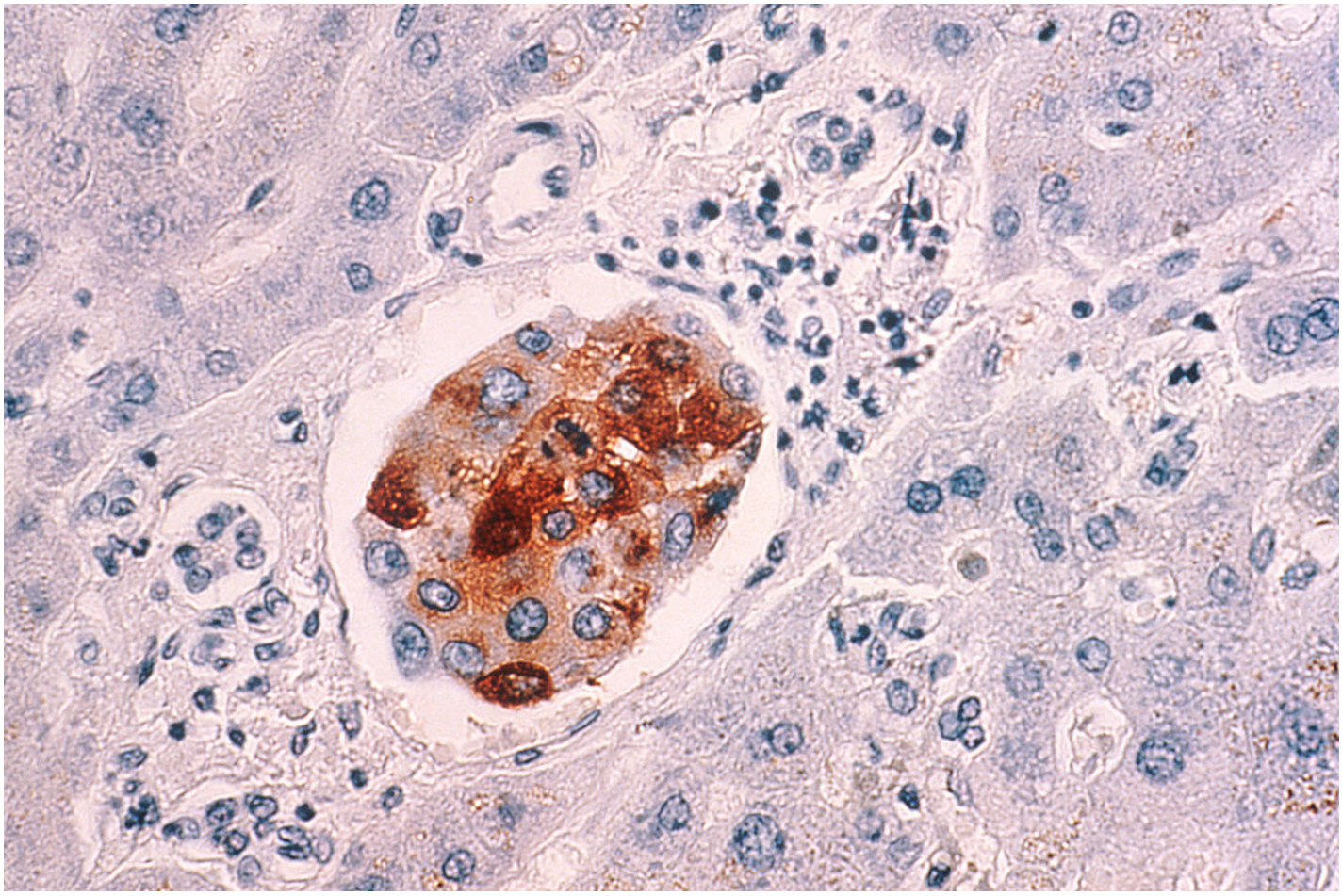
An IV bottle used in chemotherapy.
(Source: Wikimedia, National Cancer Institute; Photo credit: Linda Bartlett)

would allow for the growth of these cells were activated. Seeing these results, they sought to investigate if these backup systems were also activated in treatment-resistant cancer cells. By performing various studies in melanoma and triple-negative breast cancer cells, they saw that when these cells were treated with kinase inhibitors, the backup systems seen in endometrial cells were activated. Since the activation of the backup system depends on the presence of specific proteases,the researchers also demonstrated that there was a correlation between the effectiveness of treatment with a kinase inhibitor and the amount of these proteases present in blood samples of patients. Fortunately, the researchers say that these backup systems can be circumvented if the patient takes an additional drug that knocks out these pathways, such as an AXL inhibitor. Studies like this one help us come closer to exposing and targeting the weaknesses of cancer.

—S. SANTIAGO

ONCOLOGY

Preventing Metastasis



Malignant breast cancer cells (with brown cytoplasm) metastasized in a human liver.
(Source: Wikimedia, Photo credit: National Cancer Institute.)

Cancer cells are able to spread from a starting tumor to other parts of the body. This process, known as metastasis, is the source of approximately 90% of all human cancer deaths. In order for cancer cells to metastasize, they must first travel through the surrounding tissue and enter a blood vessel. Frank Gertler, a MIT Professor of Biology and member of the Koch Institute for Integrative Cancer Research, believes that by blocking the cancer cells from travelling through the tissue, the cells become nonmetastatic.

Gertler, along with a research team including postdoc Madeleine Oudin and 13 other scientists affiliated with the Koch Institute, found that cancer cells move along an environmental gradient from low to high concentrations of fibronectin, a protein found in the extracellular matrix surrounding the edges of both tumors and blood vessels. Furthermore, the group identified MenaINV, a variant of the cell migration modulating Mena

protein, as a critical component in the process.

MenaINV contains an additional segment not found in Mena which allows for the MenaINV to bind more strongly to the alpha-5-integrin receptor found on the surfaces of tumor cells for fibronectin recognition. As a result, the normally tangled fibronectin proteins bind to the receptor and stretch out into long bundles, attracting collagen and forming thick fibrils that protrude from the tumor cells, paving a path for the tumor cells to migrate along.

In studies of mice, the group found that cells with MenaINV were able to recognize and travel towards higher concentrations of fibronectin along collagen pathways, while cells with only the standard Mena variant were not able to migrate.

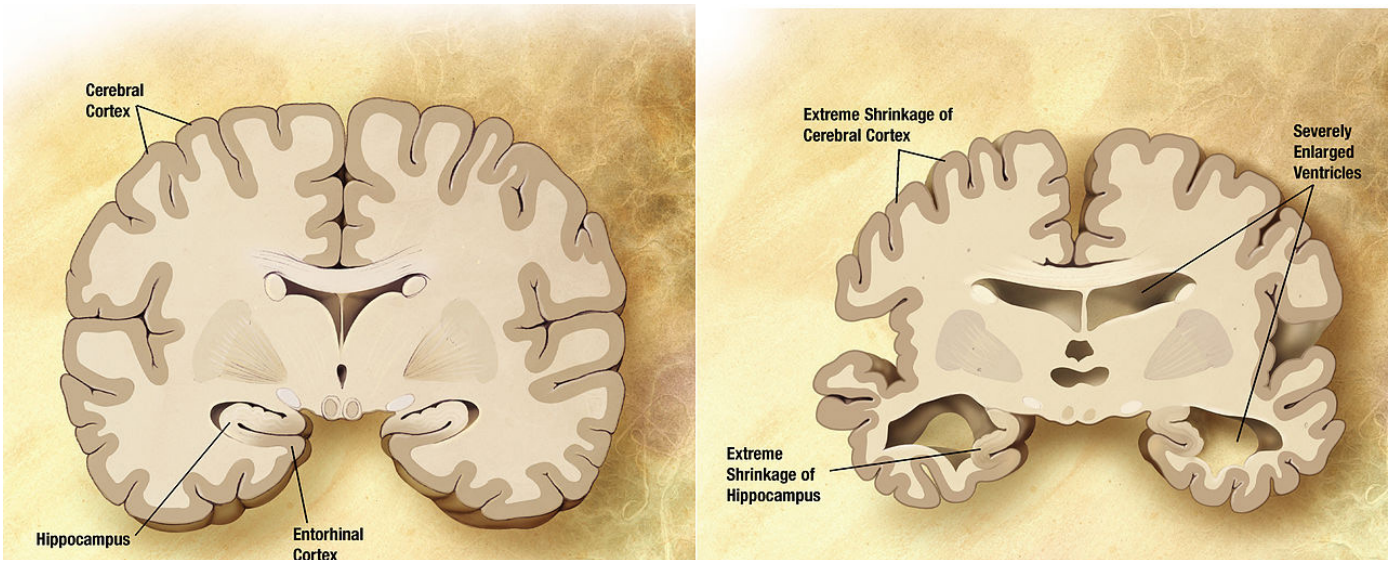
The researchers also looked into data from breast cancer patients and found that high levels of MenaINV and fibronectin correlated with metastatic tumors and cancer-related death, while no correlation was present between the standard variant of Mena and death.

Gertler and his lab previously worked on the development of antibodies that could identify Mena and MenaINV in patient biopsy samples. Such readings could help doctors predict the metastatic potential of tumors and possible options for treatment. Furthermore, research can be done on the development of drugs that inhibit MenaINV, effectively preventing tumors from becoming metastatic. Researchers are also investigating the role of MenaINV in other cancers, the mechanism behind the production of Mena and MenaINV, and the role of the other extracellular matrix proteins in metastasis.

—T. Alam

NEUROSCIENCE

Retrieving Lost Memories



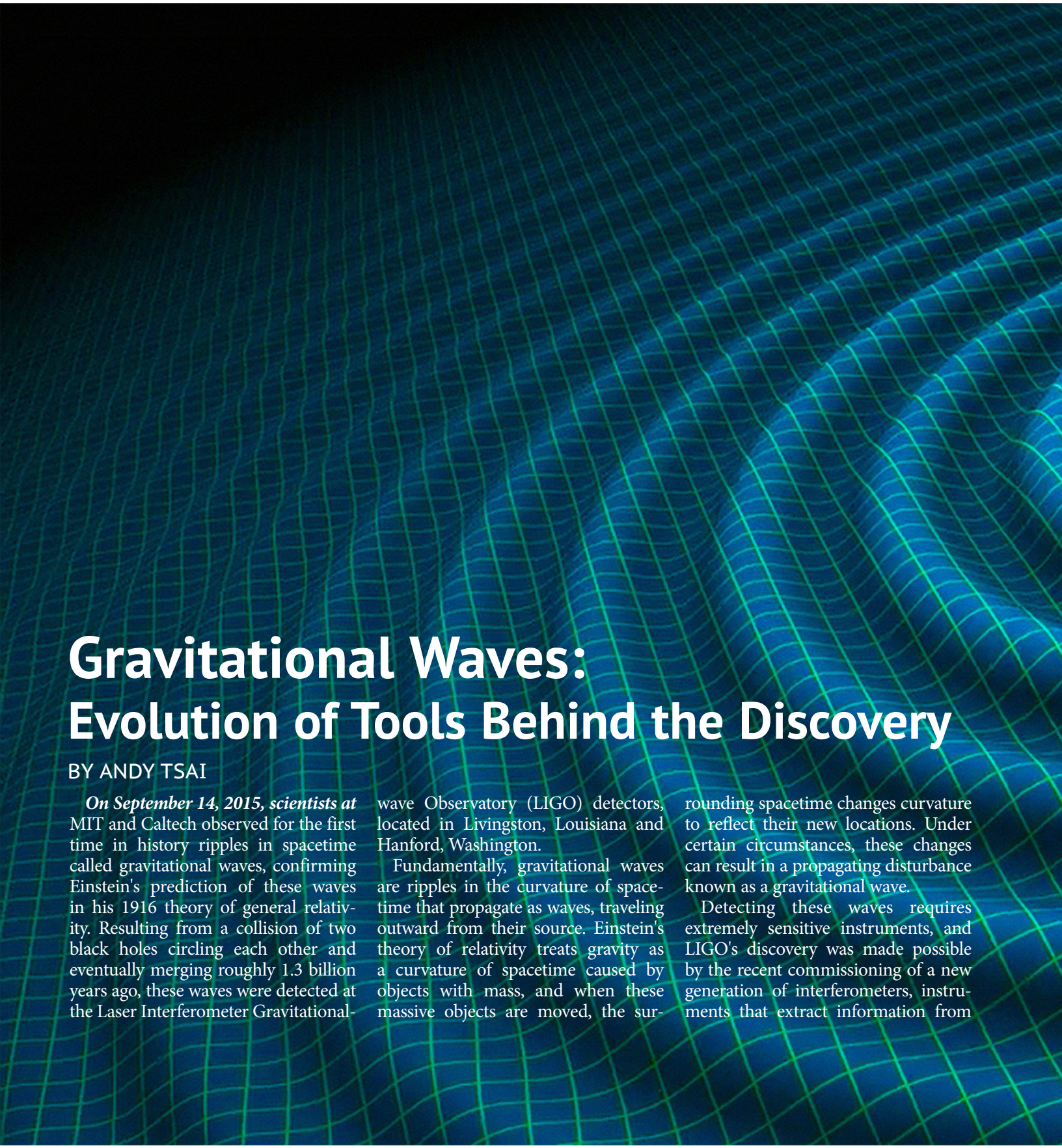
Comparison of healthy brain (left) to the brain of an Alzheimer's patient (right). (Source: Wikimedia)

Alzheimer's disease is a neurodegenerative disease that affects around 48 million people worldwide. A symptom of the early stages of Alzheimer's, and perhaps its most defining one, is short term memory loss. Fortunately, according to a paper published in the journal Nature by MIT researchers from the Tonegawa lab, these memories may not actually be "lost".

Research in the Tonegawa lab is focused on unraveling the mechanisms that underlie learning and memory. After observing that mice with retrograde amnesia were able generate new memories, the scientists sought to investigate whether or not mice suffering from early-stage Alzheimer could also form new memories. Following a preliminary study with mice, the researchers saw that the short-term memory of mice with early-stage Alzheimer's was functional in the scale of hours, but long-term memory was impaired. Then, using optogenetics, the researchers showed that the reason behind the impairment was not because of the deletion of these memories, but rather due to the inaccessibility of newly formed memories to mice. However, as the scientists also showed, these memories can be made accessible once again by using optogenetics to stimulate the formation of neural connections between the hippocampus and the entorhinal cortex, which are both involved in memory. Although the approach has its limitations, it is clear that it paves the way for the creation of a treatment that can be used to revert memory loss in Alzheimer's disease.

—S. SANTIAGO

MURJ Features



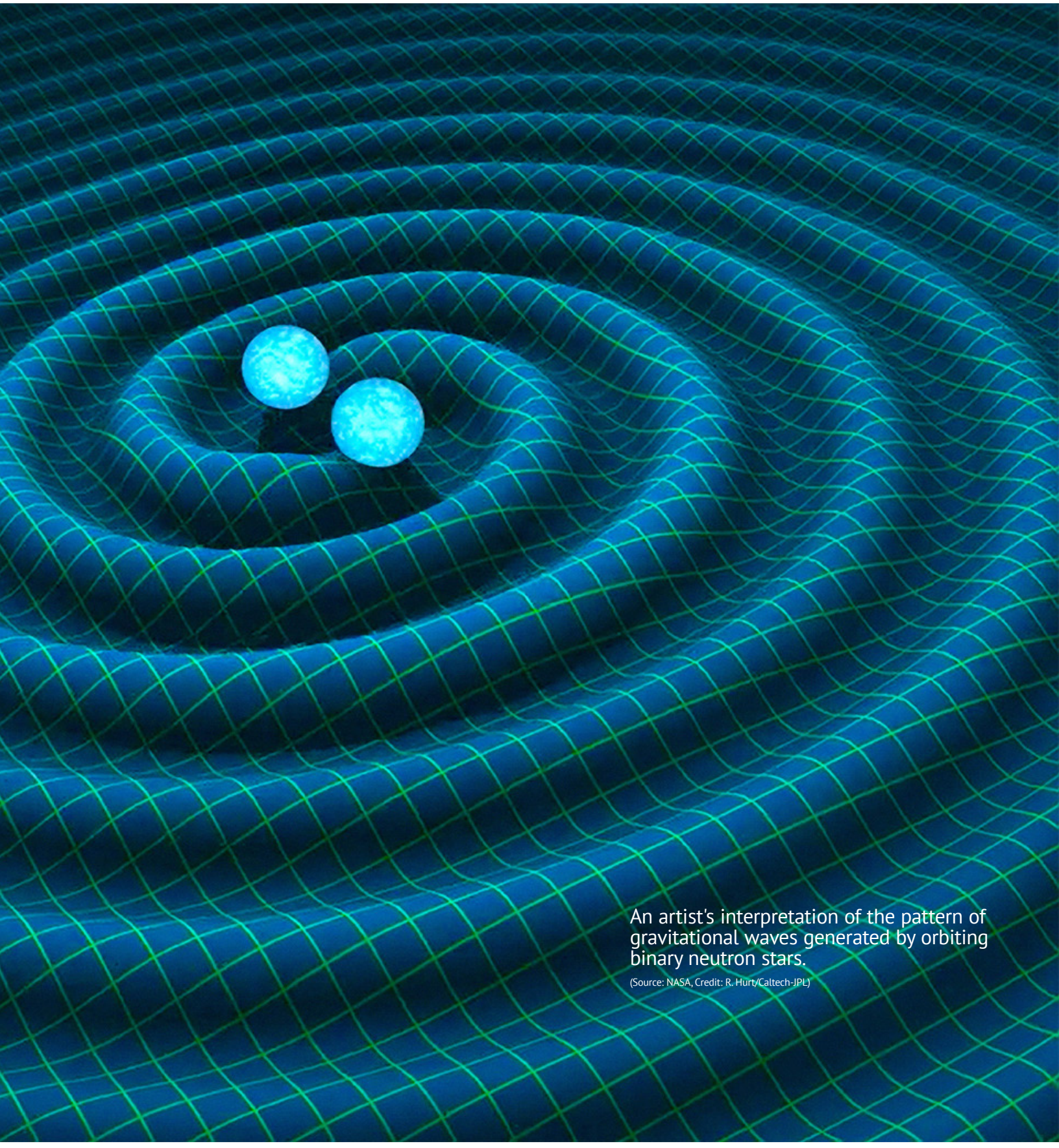
Gravitational Waves: Evolution of Tools Behind the Discovery

BY ANDY TSAI

On September 14, 2015, scientists at MIT and Caltech observed for the first time in history ripples in spacetime called gravitational waves, confirming Einstein's prediction of these waves in his 1916 theory of general relativity. Resulting from a collision of two black holes circling each other and eventually merging roughly 1.3 billion years ago, these waves were detected at the Laser Interferometer Gravitational-wave Observatory (LIGO) detectors, located in Livingston, Louisiana and Hanford, Washington.

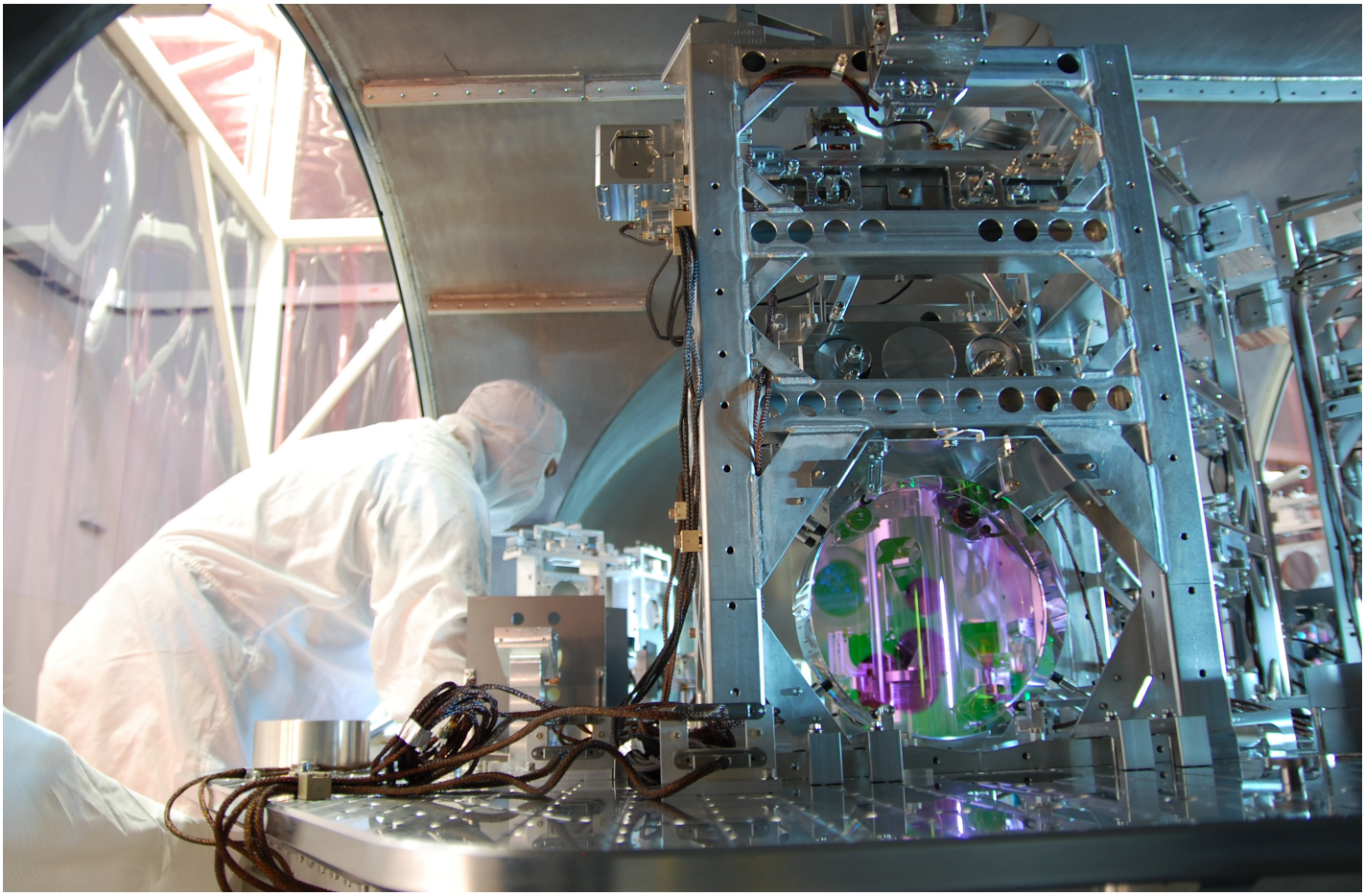
Fundamentally, gravitational waves are ripples in the curvature of spacetime that propagate as waves, traveling outward from their source. Einstein's theory of relativity treats gravity as a curvature of spacetime caused by objects with mass, and when these massive objects are moved, the surrounding spacetime changes curvature to reflect their new locations. Under certain circumstances, these changes can result in a propagating disturbance known as a gravitational wave.

Detecting these waves requires extremely sensitive instruments, and LIGO's discovery was made possible by the recent commissioning of a new generation of interferometers, instruments that extract information from



An artist's interpretation of the pattern of gravitational waves generated by orbiting binary neutron stars.

(Source: NASA, Credit: R. Hurt/Caltech-JPL)



A technician works on one of LIGO's optics. At each observatory, the 2.5-mile long L-shaped LIGO interferometer uses laser light split into two beams that travel back and forth down the arms. (Source: NSF, Photo Credit: LIGO Laboratory)

measurements of the way light interacts. By measuring the way light and other types of waves interact with objects in space, scientists can gain information about the universe. The sensitivity of an interferometer is crucial to its performance, so 'commissioning'—defined by LIGO as "the process of improving the instrument systems and computing infrastructure of the LIGO interferometers"—of the tool is key.

"The real breakthrough was the design of second-generation interferometers, also known as Advanced LIGO, that we knew would be much more sensitive," said Dr. Peter Fritschel, a professor at MIT's Kavli Institute working in the LIGO collaboration.

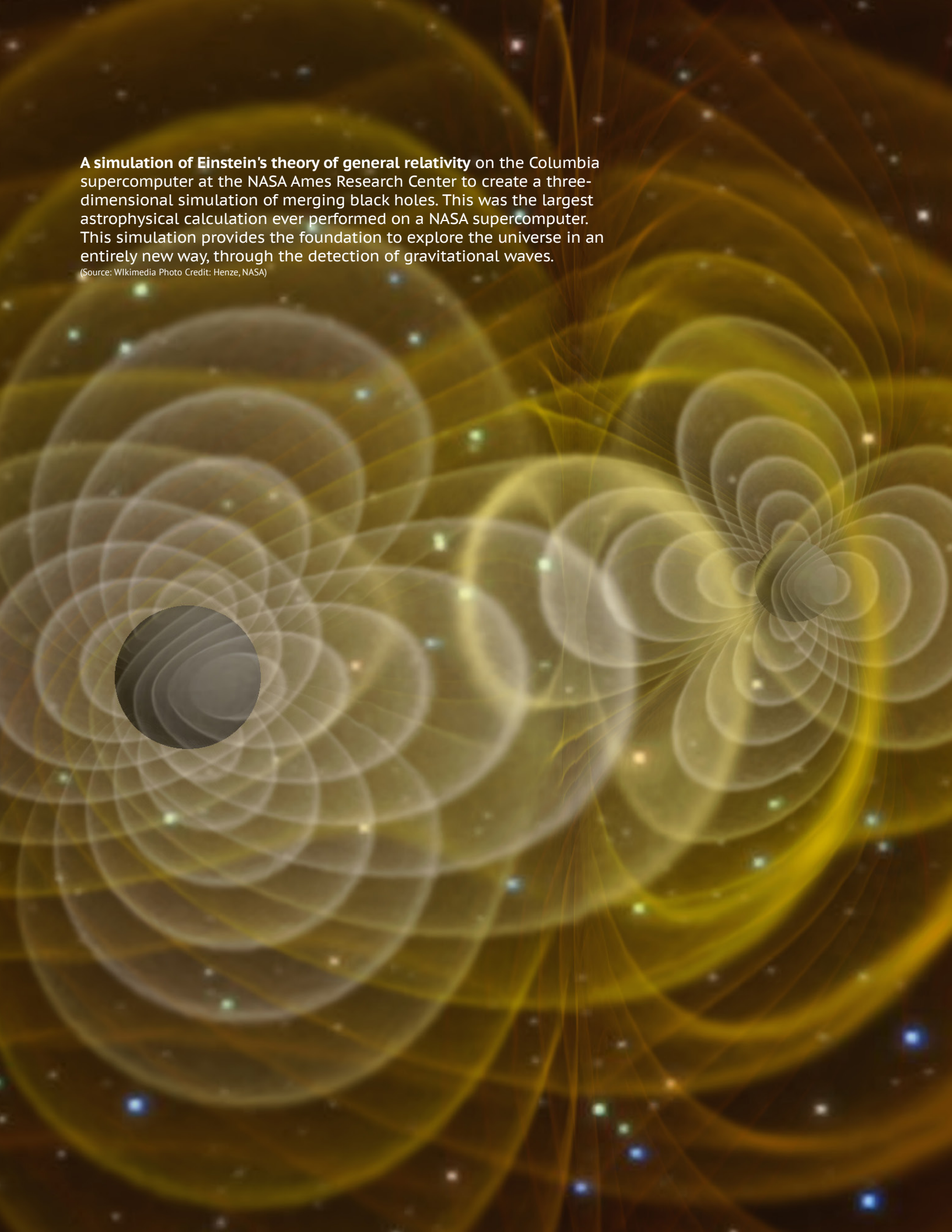
Though LIGO's interferometers have been around for decades, it was not until the most recent commission of these instruments that the detection of gravitational waves occurred. In fact, the design and commissioning of these instruments had been years in the making. While scientists were operating first-generation interferometers in the early 2000s, researchers were already simultaneously designing the Advanced LIGO, which replaced old interferometer parts with new improved ones: lasers with more powerful lasers, old test masses with new test masses, and old ground vibration isolation systems with new ones.

"The real breakthrough was the design of second-generation interferometers."

Around the end of 2010, the operation of the first-generation interferometers was terminated, and the commissioning of second-generation interferometers began.

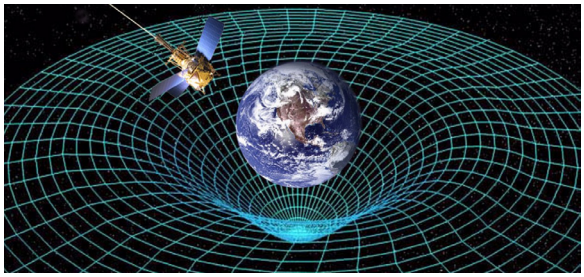
"We knew that everything was designed to be more sensitive in Advanced LIGO while we were operating the first generation interferometers," reflected Dr. Fritschel. "In the design of second-generation interferometers, we put in a lot of new features from a lot of things we learned form the first-generation- new features that made them much easier to operate."

The commissioning of second-generation interferometers used a staged approach. For six months at a time, scientists and engineers would fine-tune the sensitivity of these instruments, and once they'd reached a certain goal, Advanced LIGO was opened to take



A simulation of Einstein's theory of general relativity on the Columbia supercomputer at the NASA Ames Research Center to create a three-dimensional simulation of merging black holes. This was the largest astrophysical calculation ever performed on a NASA supercomputer. This simulation provides the foundation to explore the universe in an entirely new way, through the detection of gravitational waves.
(Source: WIkimedia Photo Credit: Henze, NASA)

Another output from a simulation run on the Columbia supercomputer.
(Source: WIkimedia Photo Credit: Henze, NASA)



Above: A depiction of spacetime being warped by the Earth's mass.

(Public Domain)

At Right: An aerial view of the LIGO detector in Livingston, Louisiana.

(Source: NSF, Photo credit: LIGO Laboratory)

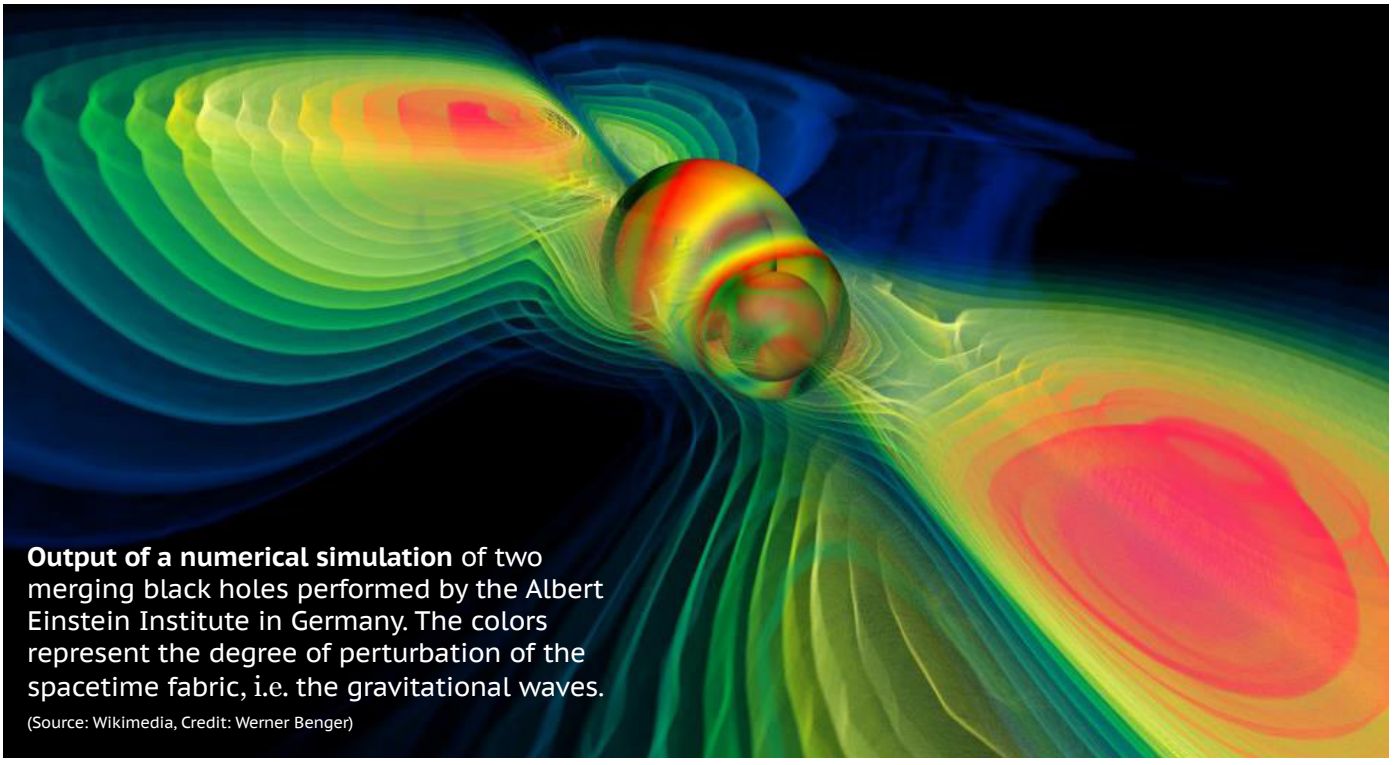
data for another six months. In its most recent commissioning period at the end of November 2014 to August 2015, the second generation interferometers were installed in both detectors of LIGO.

"We're working on making them work properly, [and] getting them sensitive enough. It was a lot of work to install the instruments at a sensitive level," Dr. Fritschel said.

For advanced LIGO, the sensitivity goal was reached in August, and the interferometer officially opened for data collection in September.

"The most fulfilling part of the journey was certainly the detection. We never expected to detect gravitational waves so early into our data collection," Dr. Fritschel said. "It would have been three decades since I started working on gravitational waves, and we were lucky to have the detection of the waves be an event as interesting as the collision of two black holes."

As of now, Advanced LIGO is back in the commissioning stage. The instrument is projected to be 2.5 to 3 times more sensitive in its full capacity, and these upgrades will slowly be implemented in the next few years. ■



Output of a numerical simulation of two merging black holes performed by the Albert Einstein Institute in Germany. The colors represent the degree of perturbation of the spacetime fabric, i.e. the gravitational waves.

(Source: Wikimedia, Credit: Werner Benger)

MURJ Spotlight: MIT Department of Biology Hidde Ploegh

This issue's spotlight features Professor Hidde Ploegh, a member of the Whitehead Institute and a Professor of Biology at MIT.

BY RACHEL ROCK

Just as lymphocytes travel the body in search of pathogens, immunologist Hidde Ploegh has traveled the world in pursuit of his research interests. A native of the Netherlands who received a Master of Science degree in biology and chemistry in 1977 from the University of Groningen, Ploegh has taught at the University of Cologne, Harvard Medical School, and MIT. Though Hidde Ploegh's multitude of achievements—from serving as a correspondent of the Royal Dutch Academy of Sciences to winning the Interbrew-Baillet Latour Health Prize, the Havinga Medal, the Avery-Landsteiner prize and a National Institute of Health Director's Pioneer Award—are nearly endless, they can be quantified. His immense curiosity, however, cannot. Ploegh has not only vitally contributed to our understanding of antigen processing and the mechanisms by which viruses evade the immune system, but also has never lost his boundless curiosity. Currently, Ploegh's lab is exploring next-generation chemical tools including sortases, or bacterially derived enzymes involved in the transfer of acyl groups. Ploegh's research is a key driving force towards the future of immunotherapy.

The field of immunology has been evolving rapidly and many of the greatest discoveries have occurred during your lifetime, so how have discoveries such as the structure of an antibody molecule and the evolution of dendritic cells caused you to evolve as a scientist as well?

I'd say the discovery of the antibody molecule is a relatively ancient discovery and we pretty much take it for granted, so I think the most important thing to remember is the antibody technology that emerged mid-seventies revolutionized the field. From the year 2000 onwards, the ability to manipulate antibody structure by genetic engineering has massively expanded the things one can do with antibodies.

We're particularly enthused about the discovery in 1993 by Belgian investigators that camelids not only make the classical type antibodies that we've come to appreciate but they also make antibodies composed of heavy chains only. This affords, yet again, new possibilities for genetic engineering, for creating antibody-based tools with properties that can be harnessed both for diagnostic applications and possibly for therapy.

Dendritic cells are a different story. They were discovered in the 70s if my memory doesn't fail me. They're now perceived as the orchestrators of what we call the adaptive immune response. So, as an antigen—a foreign substance—enters the body, it's commonly held that

Professor Hidde Ploegh is advancing key research in immunology, exploring next-generation chemical tools such as sortases.
(Photo credit: Ploegh Lab archives)

no real robust response can ensue unless this antigen is handled by professional antigen-presenting cells, and, among those, dendritic cells reign supreme.

So, can you tell me more about the research you've done with dendritic cells and MHCs? Just as the immune system hones in on antigens, a lot of your career, especially your early career prior to the work you've done with sortases, has been really targeted at MHCs. What first interested you in them and what kept your interest going so long?

Well, my thesis advisor, along with the late Don Wiley, discovered, I think, the molecular mechanism of antigen presentation through a structural biology approach. They purified MHC products, crystallized them, and then solved the X-ray structure of the protein, and what they found, remarkably, is a mechanism that explains how short snippets of protein are presented by these MHC proteins, such that T-cells can recognize them. It's now clear that, again, dendritic cells play a key role in acquisition of antigen and converting the acquired materials into protein fragments that can be presented by these MHC products. And that's ultimately what the T-cell sees and what it will recognize, for example, in a virus-infected cell.

Now, what we didn't know at the time I got involved in this field was the detailed biochemical and cell biological mechanisms by which these conversions occur. We didn't really understand how MHC products were put together from their constituent subunits. We didn't really understand the pathway traveled by these MHC products within the cell. And, because of the different functional distinctions between MHC products, it became obvious pretty quickly that -not based on our results, but based on what other immunologists had told us- that the two major classes of MHC product, class I and class II, would sample mostly distinct cellular compartments, and this became a very interesting cell biological question: How do we put together glycoproteins from their constituent subunits? How do we ensure that they travel to the right destination and pick up protein fragments? How do we ensure that once these complexes have formed, they're displayed in a way that the T-cell can recognize them?

>500
The number of papers Ploegh has contributed to

"I find MIT a quintessentially curiosity-driven environment, which I like."

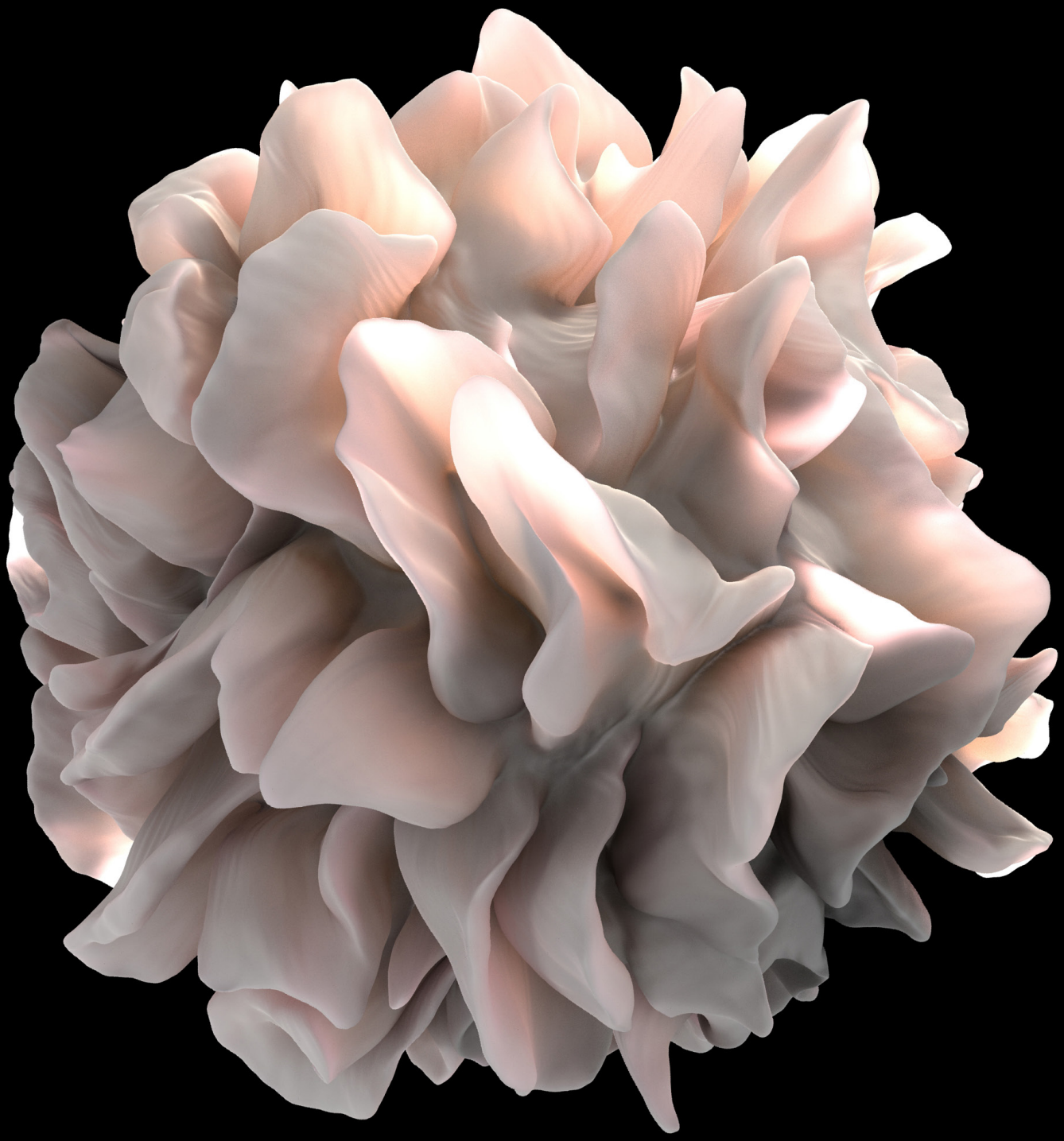
Then, part of that logic includes the notion that the immune system evolved to defend us against invaders like viruses. Given the short replicative time required to make the next generation of virus, it's easy for the virus to generate a number of mutants that allow it to escape from immune recognition.

Now, different families of viruses do this in different ways. HIV does this by simply peppering the genome with mutations so that the protein fragments generated may not be recognized properly by the available T-cells. The rate of mutation is so high that the immune response almost always lags behind the evolutionary capacity of the virus, so you create far more new variants than the speed with which your immune system can respond to them. The influenza virus does something conceptually similar. If you get the flu and recover because your immune system recognizes the virus, the immune response you mount includes antibodies and these antibodies impede binding of virus to susceptible cells—they may inhibit fusion of the virus envelope to the target cell, they may do all sorts of different things. But, of course, any virus that sustains a mutation at the site recognized by an antibody will no longer be attacked by the immune system, so there is a strong selective pressure exerted by the immune system itself on the virus such that escape variants are selected and enhanced. So, HIV and flu I would put more or less in that category of rapidly being able to create a number of mutants that can escape immune recognition.

The large viruses like herpesviruses and poxviruses work very differently. Work that I did when I first came to MIT focused on the mechanism by which herpesviruses in particular managed to elude immune detection. The way in which they do this is not fully understood, but they make a family of small glycoproteins, some of which selectively target these MHC products for degradation, and that has given us some interesting cell-biological insights into aspects of glycoprotein quality control. This led me to explore aspects of cytosolic proteolysis by proteasomes as well as the ubiquitin system and also allowed me to explore some chemical approaches to design new tools for mixed interference processes.

Can you tell me more about some of those chemical approaches?

Many of the steps involved in cytosolic proteolysis can be blocked by short synthetic contents that have been suitably modified. Basically, you create a short peptide that can be recognized by the protease involved. You can then equip this short peptide with a chemical warhead that hits the enzyme's active site and kills it and so these become what are known as active-site directed inhibitors. And so, lucky for us, a key protein involved in cytosolic proteolysis is called the proteasome—it's a huge molecular machine. A graduate student in the lab, Matt Boygo, succeeded in developing novel classes of peptide-based inhibitors of that protease and that allowed us to demonstrate involvement of proteases in the processes I've just described.



Artistic rendering of a human dendritic cell illustrating sheet-like processes that fold back onto the membrane surface. When exposed to HIV, these sheets entrap viruses in the vicinity and focus them to contact zones with T-cells targeted for infection.

(Source: National Cancer Institute, Creators: Don Bliss, Sriram Subramaniam)

Okay. On another note, you've really travelled around quite a bit throughout your career. You've taught at the University of Cologne, you've taught at Harvard. What's drawn you to MIT, specifically?

I was first recruited to MIT by my colleagues at the Center for Cancer Research. Susumu Tonegawa was there at the time. Richard Hynes was its director. It was just an exciting place to be. Coming from Europe and to be able to join the MIT faculty was a dream come true. I've always been interested in teaching as well as research and so the opportunity arose for me to assume the leadership position of the Graduate Program of Immunology at Harvard Medical School, which I then did for the next eight years. Then, I returned to MIT because I felt that MIT was more of a curiosity-driven place than Harvard Medical School, which is not to say anything bad about Harvard Medical School. It's just the flavor of how people approach problems. A medical school, understandably, is focused on issues of relevance to medicine—diagnosis and treatment of disease, understanding the molecular basis of disease and so forth—whereas MIT, with its engineering departments, robotics, artificial intelligence, chemistry, and physics, offers, I'd say, a much broader palette of disciplines, all of which can contribute to the scientific ambiance. I find MIT a quintessentially curiosity-driven environment, which I like.

You're certainly a very curious, passionate, and involved person. Your research spans a variety of areas in immunology, and you've contributed to over four hundred papers—

Five hundred.

Wow! That's amazing! So, how does your curiosity manifest for you both in research and also outside of research?

I think it's always nice to think about new ways of approaching existing problems and this often requires the development of new technology, and I've learned that that's something I really enjoy. The use of these peptide-based active-site inhibitors is one example where we sort of developed a class of these compounds. This was then extended to inhibitors based on the ubiquitin protein itself. We discovered inhibitors for other enzymes involved in this degradative pathway. The sortase approach to which you referred is an example of a protein modification technique that has proven to be quite versatile, especially in conjunction with the unique classes of antibody made by the camelids—alpacas and so forth. I've always been intrigued by putting different elements together in a novel way and seeing what emerges at the other end.

A current fascination is the antibody fragments derived from camelids. It turns out these can be used for a variety of purposes. They serve as crystallization chaperones, which allows you to infer structural details that may otherwise be hard to get. They can be used in imaging approaches, positron emission tomography. They can be used to block enzymatic interactions inside cells. We've learned to appreciate them as a really versatile, novel tool, and especially in conjunction with these protein modification methods, this becomes quite powerful.

With regard to current research, I know you're not a CRISPR/Cas9 lab, but I recently read a paper from your lab in regard to CRISPR/Cas9 and increasing the efficiency of this technique. Can you tell me more about this?

This is really to the credit of Takeshi Murayama, first author of this paper, and he was looking for a method with which you could improve the efficiency with which you can use Cas9/CRISPR to make the desired insertions into the genome and thinking about the molecular mechanisms involved. The inhibitor he came up with, which blocks non-homologous end-joining, promotes homologous recombination. So, for me, this is an interesting sort of footnote. It's a useful contribution, but we're not involved in any way, shape or form with pushing the technology.

With regard to all the projects you've worked on, you've won several different prizes—you've actually been named a pioneer by the National Institute of Health, you've won the Avery-Landsteiner Prize, and as we've discussed, you've done a myriad of different things in research. What accomplishment would you say you're most proud of?

I'd say what is a really satisfying experience is to see the confluence of some of these protein engineering methods with the use of the single domains that we produce. So, exploiting the alpacas, which serve as a source of these single-domain antibody fragments, and joining that at the hip with these sortase-based protein modification methods we developed, I think that allows us to do really elegant labeling experiments. Even though those papers may not be the most highly cited to come from our lab, it's something that gives me great satisfaction.

How do you see this evolving in the future?

I have no idea. I mean, you'd like to think that at some point, some of these technologies gain wider acceptance, and so we're by no means the first to use these single domains as a crystallization chaperone or these tools to elucidate underlying biology. It's nice to see that other colleagues at MIT are picking up on this and want to get in on producing tools of this type, and we try to provide assistance to make it possible. We had a very fruitful collaboration with Tom Schwartz in the biology department, we're collaborating with Steve Bell in the biology department, and we're collaborating with Richard Hynes in the Koch Center, all of whom have expressed an interest in using these single domains as a research tool. I think by having other colleagues work on different models, get in on the game, we will learn more about the mutations and possibilities of this new technology.

So, how do you see immunology as a whole evolving?

Well, 2014, Science labeled immunotherapy the breakthrough of the year, and that's really something to be keeping a close watch on. For some cancers where chemotherapy was no longer effective and surgical options had been exhausted, clever immunologists had discovered other means of turning one's own immune system against the cancer. I'd say there's now compelling evidence for cancers like melanoma and non-small-cell lung cancer that these forms of what we call immunotherapies really make a difference. Patients who would have died otherwise are now given an extended lifespan, sometimes people even seem to be cured—patients five years, tumor-free. For a cancer type like metastatic melanoma, I'd say it's a towering moment. So, there's a lot of excitement about learning how to exploit monoclonal antibodies for therapeutic use, and we've learned enough about T-cells to begin to use those as entities that can fight cancer, notably, blood-borne cancer, and so the whole field, I think, is still accelerating as we learn more about the possibilities of immunotherapy. Things will continue to improve, combining immunotherapy with other forms of treatment to improve efficiency. New targets will be discovered. I see the next decade there being very significant progress on how we exploit our immune systems to fight diseases.

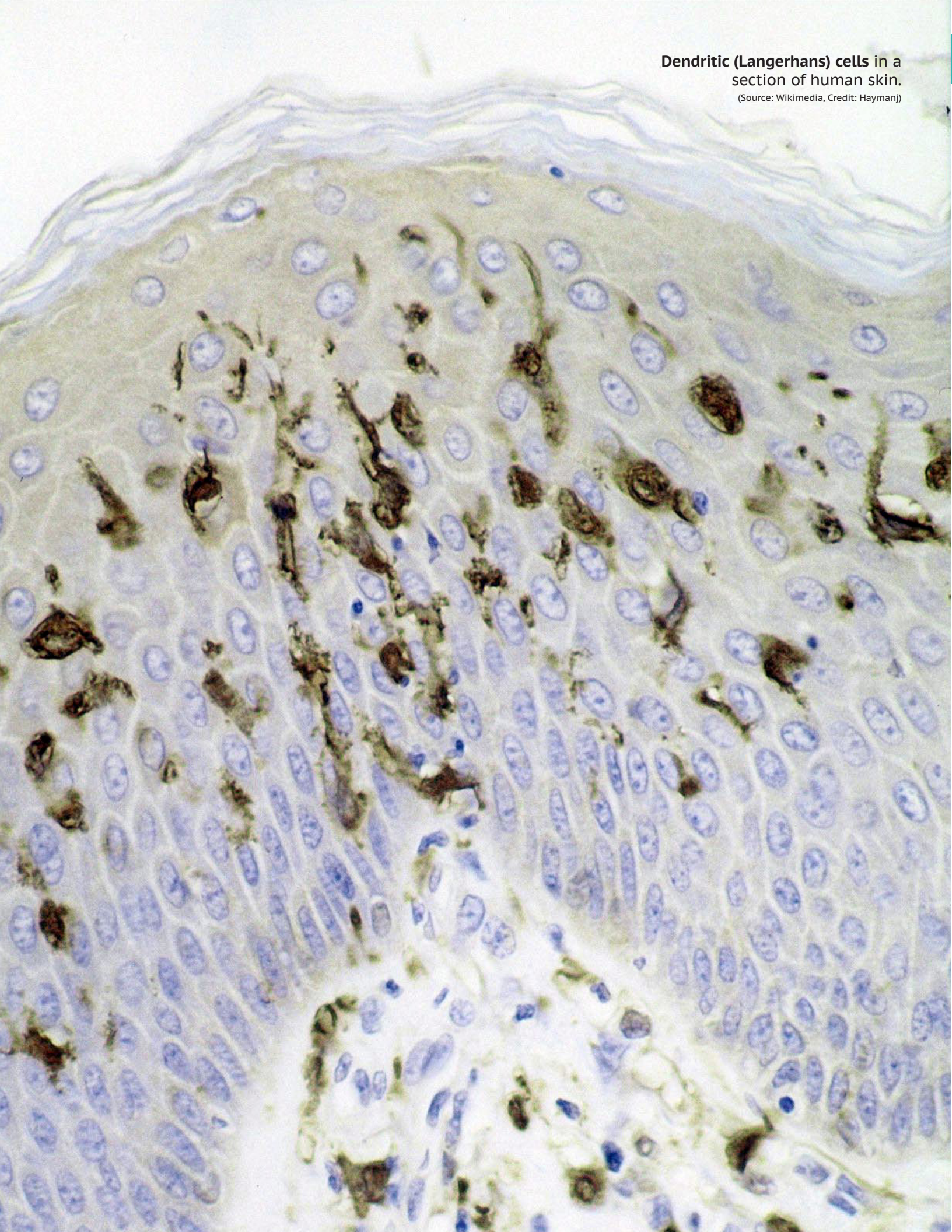
What excites you the most?

No one thing in particular, I'd say any new result that emerges, expected or unexpected, makes you think about what could have caused that, and if the result comes out in accord with your predictions, you'd say that, "My understanding of immunobiology must be at least partially correct if not entirely correct. If you get an unexpected result and this repeats itself, you'd ask, "Well, my original working model needs readjustment and so can I, in spite of these new, unexpected data, design an experiment closer to what I believe is the underlying mechanism?" So, it's any result, large or small, that tickles one's curiosity.

Members of Ploegh lab enjoy a scenic natural outing.
(Ploegh Lab)

Speaking of tickling one's curiosity, trial and error, and going through a journey—you've certainly experienced an amazing journey throughout your own life—there are many freshmen who've just been accepted to MIT on Pi day, so is there anything you would want to say to them in regard to their curiosity that you think is important for a young person in science to know?

Don't stifle your curiosity. Don't get jaded by things that don't work entirely as expected. At MIT, I would say faculty may well be among the most underutilized resources available to students. I think undergraduate students, be they junior or senior, should not be reluctant to approach faculty, because I believe we work in a research university with undergraduates as we see that as part of our task. Most of my colleagues really enjoy interacting with students at all levels. Students don't come with preconceived ideas about how things should work and they may have insights long before us who have been in the field longer because we've always assumed certain things to be the way they are, and sometimes it takes taking a fresh look at an old problem to break that jam. So, I'd say, don't be afraid to express your points of view or your ideas, don't be reluctant to contact faculty if you have something to offer, and enjoy what you're doing. If you're not enjoying what you're doing, then something's probably not right. ■



Dendritic (Langerhans) cells in a section of human skin.
(Source: Wikimedia, Credit: Haymanj)

MURJ UROP Summaries

Coordination Strategies for Effective Human-Robot Team Communication

Niyati Desai[1], Abhizna Butchibabu[2], Julie A. Shah[3]

[1]Student Contributor, Class of 2019, Computer Science & Artificial Intelligence Laboratory (CSAIL), Department of Physics, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

[2]Student Contributor, Ph.D. Candidate, Computer Science & Artificial Intelligence Laboratory (CSAIL), Department of Aerospace Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

[3]Professor, Computer Science & Artificial Intelligence Laboratory (CSAIL), Department of Aerospace Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

Abstract

The problem of effective human-agent communication is attracting attention as autonomous systems gain prevalence. We developed an autonomous agent communication model designed to emulate communication strategies used effectively by humans to achieve high levels of team coordination. This model was a Maximum Entropy Markov Model (MEMM), designed based on human teams' communication data; it yielded a prediction performance accuracy of 92.8%. We successfully implemented the model, integrated it with human teams and analyzed the teams' performances. In addition, our empirical study showed that the human-agent teams using this model had equivalent team performance to that of the human-only teams.

Background

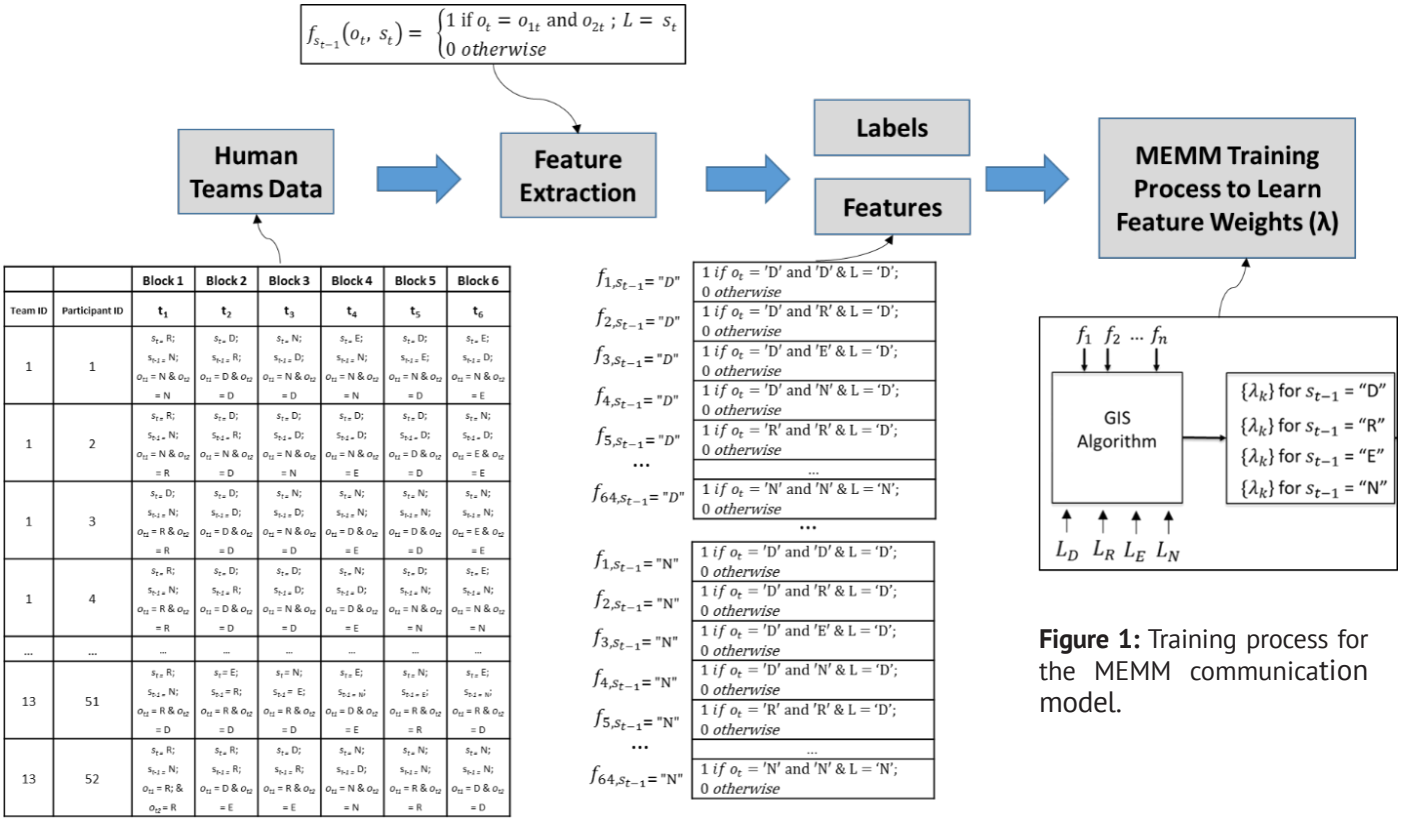
The effective integration of robots into human teams is drawing attention in many applications, especially where complex and safety-critical tasks must be performed. We drew inspiration from human teams that demonstrated effective team coordination by sharing information before it is needed (implicitly) as opposed to explicitly requesting teammates to perform actions or exchange information (Entin, Serfaty, 1998, Butchibabu, Sparano-Huiban, Shah, 2016). With the objective of ultimately preserving performance and decision-making within human-human teams, our goal was to develop a computational model for autonomous agents to communicate with human teammates. We compared the performance of human-agent hybrid teams for three different previously observed communication-type models implemented by the autonomous agent: reactive-implicit communication, deliberative-implicit communication, and MEMM (our model). Deliberative communications contain information related to the next-goal in sequence and reactive communications are status updates regarding the environment. The MEMM model is a model that has demonstrated superior performance when learning from observations compared to other commonly used learning from demonstration (LfD) models such as Hidden Markov Models (McCallum, Freitag, Pereira, 2000). Previous work led us to hypothesize that the deliberative model would result in higher performance than the reactive model and our hope was that the MEMM model would achieve equivalent or higher performance compared to that of the deliberative (Butchibabu, et. al, 2016, Shah, Brezeal, 2008). By conducting a human-subject study with teams of two agents and two humans, we were able to successfully test the three computational models and determine how overall team performance varied between them.

The human study task required the subjects to pick up and deposit colored blocks in a virtual platform in a designated sequence. The environment consisted of several rooms containing scattered colored blocks in each. The team members had to travel to each room, look inside and find the right colored blocks to deposit them in one room (the dropzone) in the right sequence. The subjects worked together as a team by exchanging communications like: which color block they are currently searching for, when they dropped off a block, if they know the room location of a specific color block for later in the sequence... etc.

The communication model we wrote aimed to allow an agent to communicate with other teammates to accomplish some complicated task. The goal was to make a model that could be smoothly integrated to boost coordination and thus decrease task completion time (measured as high performance).

Methods

We developed and tested three independent computational communication models: reactive, deliberative and MEMM. The reactive and deliberative models were implemented by simply restricting the possible communications available to the agent to only reactive and only deliberative respectively, and then choosing the appropriate communication message based on the physical scenario of the agent. The MEMM model took into account two input variables: the agent's last communication and the two immediate previous communications by others on the team. Feature functions are binary functions which represent data by depending on tuples of observations, previous state st-1 and their state labels st (McCallum et. al., 2000).



Team ID	Participant ID	Block 1 t_1	Block 2 t_2	Block 3 t_3	Block 4 t_4	Block 5 t_5	Block 6 t_6
1	1	s_{t_1} R; s_{t_2} N; o_{1t} = N & o_{1t} = N	s_{t_1} D; s_{t_2} R; o_{1t} = D & o_{1t} = D	s_{t_1} N; s_{t_2} D; o_{1t} = N & o_{1t} = D	s_{t_1} E; s_{t_2} N; o_{1t} = N & o_{1t} = N	s_{t_1} D; s_{t_2} E; o_{1t} = N & o_{1t} = D	s_{t_1} E; s_{t_2} D; o_{1t} = N & o_{1t} = E
1	2	s_{t_1} R; s_{t_2} N; o_{1t} = N & o_{1t} = R	s_{t_1} D; s_{t_2} R; o_{1t} = N & o_{1t} = D	s_{t_1} D; s_{t_2} D; o_{1t} = N & o_{1t} = N	s_{t_1} D; s_{t_2} D; o_{1t} = N & o_{1t} = E	s_{t_1} D; s_{t_2} D; o_{1t} = D & o_{1t} = D	s_{t_1} N; s_{t_2} D; o_{1t} = E & o_{1t} = E
1	3	s_{t_1} D; s_{t_2} N; o_{1t} = R & o_{1t} = R	s_{t_1} D; s_{t_2} D; o_{1t} = D & o_{1t} = D	s_{t_1} N; s_{t_2} D; o_{1t} = N & o_{1t} = D	s_{t_1} N; s_{t_2} N; o_{1t} = D & o_{1t} = E	s_{t_1} N; s_{t_2} N; o_{1t} = D & o_{1t} = D	s_{t_1} N; s_{t_2} N; o_{1t} = E & o_{1t} = E
1	4	s_{t_1} R; s_{t_2} N; o_{1t} = R & o_{1t} = R	s_{t_1} D; s_{t_2} R; o_{1t} = D & o_{1t} = D	s_{t_1} D; s_{t_2} D; o_{1t} = N & o_{1t} = D	s_{t_1} N; s_{t_2} D; o_{1t} = D & o_{1t} = E	s_{t_1} D; s_{t_2} N; o_{1t} = N & o_{1t} = N	s_{t_1} E; s_{t_2} N; o_{1t} = N & o_{1t} = N
...	...	...	...	...	...	...	...
13	51	s_{t_1} R; s_{t_2} N; o_{1t} = R & o_{1t} = D	s_{t_1} = E; s_{t_2} = R; o_{1t} = R & o_{1t} = D	s_{t_1} = N; s_{t_2} = E; o_{1t} = R & o_{1t} = D	s_{t_1} E; s_{t_2} = N; o_{1t} = D & o_{1t} = E	s_{t_1} N; s_{t_2} = E; o_{1t} = R & o_{1t} = R	s_{t_1} E; s_{t_2} = N; o_{1t} = R & o_{1t} = D
13	52	s_{t_1} R; s_{t_2} N; o_{1t} = R; & o_{2t} = R	s_{t_1} R; s_{t_2} R; o_{1t} = D & o_{1t} = E	s_{t_1} D; s_{t_2} R; o_{1t} = R & o_{1t} = E	s_{t_1} N; s_{t_2} D; o_{1t} = N & o_{1t} = N	s_{t_1} N; s_{t_2} N; o_{1t} = R & o_{1t} = R	s_{t_1} N; s_{t_2} N; o_{1t} = D & o_{1t} = D

$f_{1,s_{t-1}}$ = "D"	1 *if* o_t = 'D' and 'D' & L = 'D'; 0 *otherwise*
$f_{2,s_{t-1}}$ = "D"	1 *if* o_t = 'D' and 'R' & L = 'D'; 0 *otherwise*
$f_{3,s_{t-1}}$ = "D"	1 *if* o_t = 'D' and 'E' & L = 'D'; 0 *otherwise*
$f_{4,s_{t-1}}$ = "D"	1 *if* o_t = 'D' and 'N' & L = 'D'; 0 *otherwise*
$f_{5,s_{t-1}}$ = "D"	1 *if* o_t = 'R' and 'R' & L = 'D'; 0 *otherwise*
...	...
$f_{64,s_{t-1}}$ = "D"	1 *if* o_t = 'N' and 'N' & L = 'N'; 0 *otherwise*
...	
$f_{1,s_{t-1}}$ = "N"	1 *if* o_t = 'D' and 'D' & L = 'D'; 0 *otherwise*
$f_{2,s_{t-1}}$ = "N"	1 *if* o_t = 'D' and 'R' & L = 'D'; 0 *otherwise*
$f_{3,s_{t-1}}$ = "N"	1 *if* o_t = 'D' and 'E' & L = 'D'; 0 *otherwise*
$f_{4,s_{t-1}}$ = "N"	1 *if* o_t = 'D' and 'N' & L = 'D'; 0 *otherwise*
$f_{5,s_{t-1}}$ = "N"	1 *if* o_t = 'R' and 'R' & L = 'D'; 0 *otherwise*
...	...
$f_{64,s_{t-1}}$ = "N"	1 *if* o_t = 'N' and 'N' & L = 'N'; 0 *otherwise*

Figure 1: Training process for the MEMM communication model.

As detailed in Figures 1 and 2, the main process behind developing the MEMM model was as follows: 1) extracting features from the human data, 2) calculating feature weights based on frequency, 3) using a Maximum Entropy Distribution to factor in the two input variables and outputting the current communication state response. The features and output response were made up of four communication types as shown in Figure 1: D- deliberative, R- reactive, E- explicit, N- none. Once a method for determining the agent's next communication was developed, the predictive model was validated based on leave-one-out cross validation. This involved training the model with all human subject data except one data point and then feeding the data point in and comparing the model's outputted result to the actual human response. All data for the MEMM model was converted into feature data points (1x64 MATLAB cell arrays since there are four communication types) based on the tuple inputs described above and then fed into a logistic regression program designed in Python. The

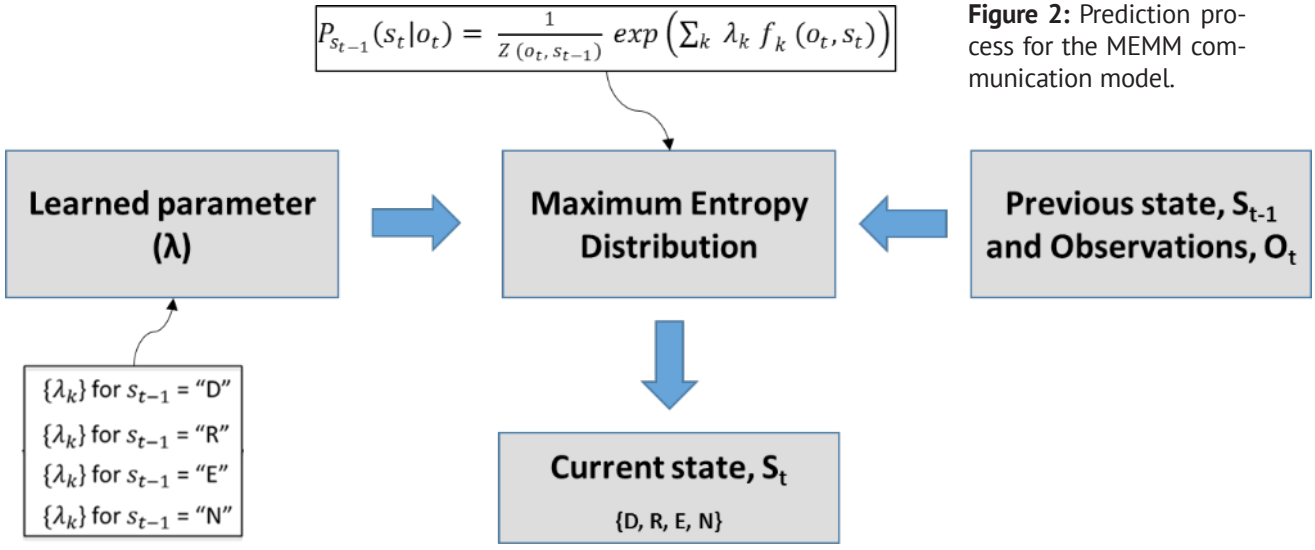


Figure 2: Prediction process for the MEMM communication model.

Python script was used to run the cross validation and after yielding satisfactory results, it was utilized to make the optimal logistic fits for each of the four st-1 communication types.

Once the fits were developed, our leave-one-out validation method yielded a performance prediction accuracy of 92.8% for the full data set (all 13 teams- 312 data points). The MEMM model was integrated into the Blocks World for Teams (BW4T) testbed which simulates a collaborative search and delivery task. Implemented in Java, BW4T has been widely used within multi-agent systems and human-robot interaction to better understand the behavior of teams involving humans and autonomous agents (Johnson, Jonker, Van Riemsdijk, Feltovich, Bradshaw, 2009). BW4T is a virtual environment where a team of agents searches for blocks and delivers them in a designated sequence. A human subject study was run to implement the model using BW4T to assess whether autonomous agents using the computational model would preserve the team performance that observed in the human teams experiment. The MEMM model used for the human subject study was only trained on the top 5 teams. Performance was primarily measured by task completion time as this has been found to be an accurate measure of effective team coordination and cooperation: lower task completion time indicated high team performance on the task (Shah et. al., 2008). With the hypothesis that the study participants would be unable to distinguish between their agent teammates and their human teammates, we also surveyed each participant on their perceptions of the human-likeness or agent-likeness of each teammate based on their communications. To remove any bias that the participants may have had towards the agent communication models, we also included a confederate participant to simulate performing the task although the actual agent active in the testbed was an agent.

Results/Discussion

Overall, our results showed that the performance of teams with agents using the MEMM communication model was statistically higher than that of teams with agents using the reactive-implicit communication and deliberative-implicit communication models (for the fastest 5 teams), defining statistical significance at the $\alpha = 0.05$ level. In addition, comparing the performance of the teams with the MEMM modeled agents to previous study's human-only team performance, we can say the mean task completion time for the top 5 teams remained equivalent within 95% confidence (Butchibabu et. al., 2016). So we can say human team performance was overall preserved with the MEMM model, but we also found a significantly larger variability in the performance of human-only teams than in performance of the human-agent teams where the agents were using MEMM models. Comparing between the three communication models, our results confirmed our hypothesis since as Figure 3 shows, the task completion time for agents using deliberative-implicit communication was significantly lower than for those using reactive-implicit communication. Analysis showed the average task completion time was significantly lower for teams with agents using the MEMM communication model than for those using the reactive-implicit only model (Figure 3). However, no statistically significant results were found when

Figure 3: Average task completion time for all 12 teams when the autonomous agents used reactive-implicit, deliberative-implicit or MEMM communication models.

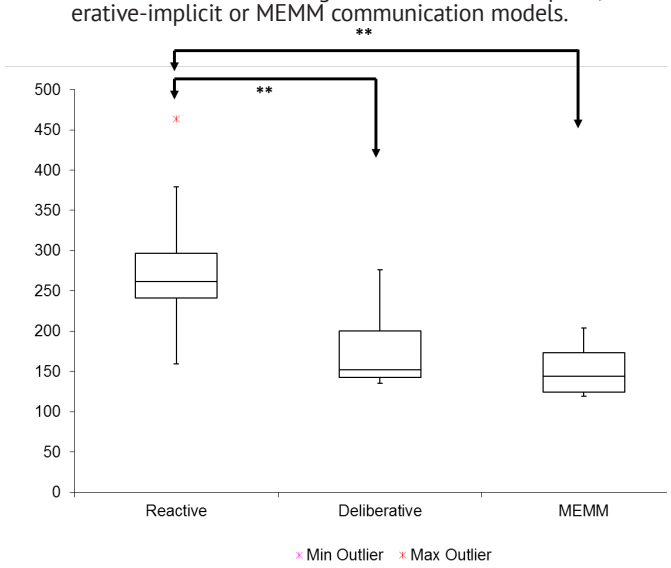
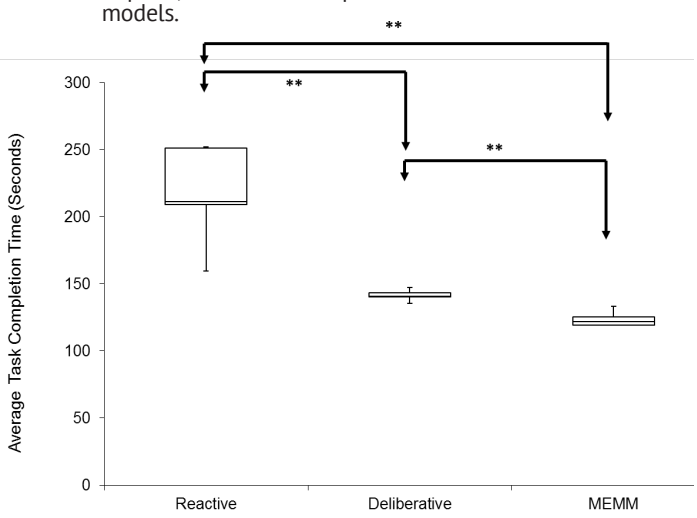


Figure 4: Average task completion time for the top 5 teams when the autonomous agents used reactive-implicit, deliberative-implicit or MEMM communication models.



comparing MEMM with deliberative-implicit communication. Furthermore, we predicted an interaction effect between communication type and team capability due to which, the top performing MEMM-modeled teams would benefit more than the deliberative-modeled teams. Further analysis (as shown in Figure 4) shows that for only the top 5 teams, not only is the performance of the deliberative-modeled teams also significantly higher than the performance of the reactive-modeled teams, but the MEMM-modeled teams demonstrated a significantly higher performance than the deliberative-modeled teams as well. Another way to identify whether the MEMM communication model was perceived to be more human-like or more agent-like than either the reactive-implicit or the deliberative-implicit communication models was to measure how accurately participants were able to distinguish their human teammates correctly. Based on the

results of the surveys conducted on participants, we were able to observe that the participants' accuracy in distinguishing their human teammates was significantly higher for the reactive model than for the deliberative model and marginally higher for the deliberative model than for the MEMM model. This indicates that the participants effectively perceived the MEMM communication model to be the most human-like.

This is one of the first studies to empirically evaluate the MEMM communication model and compare it against the deliberative-implicit communication model and the reactive-implicit communication model. This is also one of the first studies to demonstrate that an agent designed to emulate communication strategies from human teams data can also preserve human-human team communication. Overall this study has provided interesting results in refining effective autonomous agent communication modeling for human-agent teaming and merits further attention for greater avenues of future research.

References

Butchibabu, A., C. Huiban-Sparano, L. Sonenberg, and J. Shah. Evaluating Anticipatory Communication Strategies for Effective Team Coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. **2016**.

Entin, E. E., Serfaty, D., Kerrigan, C. . Choice and performance under three command and control architectures. *Proceedings of the 1998 Command and Control Research and Technology Symposium*. **1998**.

Johnson, M., Jonker, C., Van Riemsdijk, B., Feltovich, P. J., Bradshaw, J. M. Joint activity testbed: Blocks world for teams (BW4T). *Engineering Societies in the Agents World: Springer Berlin Heidelberg*.. **2009**. X, 254-256.

McCallum, A., Freitag, D., Pereira, F. C. . Maximum Entropy Markov Models for Information Extraction and Segmentation. *ICML*. **2000**. 17, 591-598.

Shah, J., Breazeal, C. An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. **2010**. 52(2)

MURJ Reports

Untraceably Proving Unique Identity with Multi-Context One-Show Credentials

Jeffrey Lim[1]

[1]Student Contributor, Class of 2015, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

This paper introduces the concept of multi-context one-show (MCOS) credentials, a cryptographic method by which a user can demonstrate possession of a credential only once within a given context, but arbitrarily many times across as many different contexts. The prototypical application for MCOS credentials is to operate cryptographically secured "polls" such that a single user can vote at most once in each poll, but the user's vote cannot be linked to their real identity or to their votes in other polls. Such credentials will be useful whenever it is required that distinct identities represent distinct persons; e.g., preventing gaming of rating- or reputation-systems, ensuring that messages originate from real humans instead of automated scripts, and enforcing norms in anonymous communities. In this paper, an MCOS protocol is proposed, which is based on the cut-and-choose method (Chaum, Fiat, and Naor, 1990) and the parallelized Schnorr protocol for zero-knowledge proofs of knowledge of discrete logarithms (Camenisch and Michels, 1999). The security properties of the protocol are examined, and its feasibility demonstrated by means of a minimal JavaScript implementation.

1. Introduction

In many cases, a participant in some communication needs to prove that they are not the same person as any of the other participants. For example, a company offering discounts to new customers would like to prevent people from exploiting the offer by signing up repeatedly under different aliases. At present, this objective is achieved simply by requiring that all participants submit their personal identifying information (drivers license, credit card number, etc.), which has adverse implications for privacy. The goal of MCOS, by contrast, is to prove one's status as a unique person while revealing no other information about oneself or one's activities.

Thus, MCOS credentials should satisfy four properties, illustrated in the abstract by the example of a poll: A user should not be able to vote more than once in a given poll ("non-duplicability"); a user's votes should not be correlatable to their real identity ("anonymity"); a user's votes in different polls should not be correlatable to each other ("unlinkability"); and the credential should not need to fix in advance a set of polls that the user can vote in ("generality"). In all of this we assume that the credential-issuer (who has the responsibility of manually verifying a user's identity) can be trusted not to issue multiple credentials to the same person, but we otherwise allow that the parties may collude to defeat the protocol.

2. Background

Prior work in the field has established various schemes that satisfy some but not all of these properties. Blind signatures (Chaum, 1983) enable a signing authority to issue signatures without knowing what they are signing, thus ensuring anonymity. In Chaum's example, a customer deposits funds at a bank and presents a blinded signature request, which the bank signs. The customer unblinds the signature to produce a signed "bill," denoting a debt from the bank to the bearer. The customer later uses this bill to purchase goods from a merchant, who presents the bill to the bank for redemption. The bank can then validate the signature as having been legitimately issued by the bank, but cannot connect it to the customer's original deposit – they know only that they issued the bill to some customer. Furthermore, these credentials are non-duplicable, since the bill is simply a piece of data whose reuse the bank can detect. (If the signature scheme is sound, the customer cannot alter the bill without invalidating the signature.)

However, Chaum's method cannot be extended to multiple polls without forgoing either unlinkability or generality. Either the voter reuses the same "bill" in every poll, making their votes easily linkable; or else they obtain a new bill for each poll, which requires that each poll be associated with a particular set of bills in advance.

Another proposal is the unlinkable multi-show credentials scheme given by Camenisch and Lysyanskaya (2001). In this system, a user can demonstrate possession of a credential arbitrarily many times via a zero-knowledge proof that prevents different uses of the credential from being connected. Thus, the user and the credential-issuer do not need to fix a set of polls in advance, and anonymity and unlinkability are maintained. However, this same property allows the user to vote an unlimited number of times in the same poll without being detected.

The protocol proposed here combines elements of both schemes to achieve all four objectives. However, it makes a slight concession in the property of generality: The user will need to obtain in advance some number of "voter-tickets" from the credential issuer, each of which will be used once and then discarded; when the user runs out of tickets, they cannot vote in any more polls until they obtain a fresh set of tickets. While this is inconvenient, it is less so than in Chaum's system, since these tickets can be freely used in any poll.

3. Description of protocol

We will now describe the protocol in terms of the interactions between a credential-issuer ("the Issuer"), a user ("Alice"), and person ("Bob") administering a poll in which Alice wants to vote.

3.1. Setup

The Issuer selects an RSA public modulus/exponent pair *(n, e)*, and a prime modulus and generator *(p, g)*, whose order is some prime q, and publishes these values.[1] (The same *p* and *g* may be reused by multiple issuers.) They keep secret the corresponding RSA private exponent *d*. All parties agree on some cryptographic hash function *H*, which takes any number of arguments. Two security parameters, *k* and *m*, are agreed upon.

3.2. Registering identity

Alice selects her secret $x \in_R \mathbb{Z}_p^*$ and finds $y = g^x \bmod p$. She sends *y* to the Issuer, along with whatever external credentials (drivers license, etc.) that the Issuer uses to adjudge the uniqueness of persons. If the Issuer is satisfied that Alice is a unique person not already in their database, they add her information and *y* value to their database. Once this has been done, Alice can obtain voter-tickets at any later time without further authentication.

3.3. Obtaining voter-tickets

Before she votes in any polls, Alice must first obtain one or more voter-tickets, each of which will be used exactly once (one ticket for each poll) before being discarded. The following illustrates the process for Alice to obtain one voter-ticket.

1. Alice selects random $r_1, r_2, \ldots r_k \in_R \mathbb{Z}_q^*$ and $b_1, b_2, \ldots b_k \in_R \mathbb{Z}_n$ (with the *b* values being relatively prime to *n*.)
2. For $1 \le i \le k$, she finds $h_i = {b_i}^e \cdot H(y^{r_i}, g^{r_i}) \bmod n$.
3. She sends all *k* of the h_i values to the Issuer, along with *y*.
4. The Issuer verifies that *y* is in their database, and chooses a random subset $C \subseteq [1, k]$, with $|C| = k/2$, and sends *C* to Alice. (Let $\overline{C}$ denote the complement of this set; i.e. $[1, k] - C$.)
5. Alice replies with (r_i, b_i) for all *i* in *C*, and the Issuer verifies that all of the corresponding h_i values fit the form specified above. (This is the "cut-and-choose" method given in Chaum, Fiat, and Naor, 1990.)
6. The Issuer replies with $S' = \prod_{j \in \overline{C}} h_j^d \bmod n$.
7. Alice finds $S = S' \cdot \prod_{j \in \overline{C}} b_j^{-1} \bmod n$. This *S*, along with the list of the remaining (y^{r_j}, g^{r_j}) pairs, comprises the voter-ticket.

In practice, Alice can significantly decrease the storage requirements by choosing $r_1, r_2, \ldots r_k$ as the deterministic output of some pseudorandom function starting from some master seed *R*; e.g. $r_i = H(R, 1)$, $r_2 = H(R, 2)$, etc. Then, she only needs to store $(S, R, \overline{C})$, which is enough for her to reconstruct the complete voter-ticket when it is needed.

3.4. Voting

Bob chooses and publishes a generator *f* (also of order *q*) with respect to the modulus *p*, which should be sufficiently random as to be unique to this particular poll, with high probability. For example, it could be produced by some pseudorandom function of Bob's domain name, timestamp, etc.[2] This *f* is known as the "poll base."

Alice finds $z = f^x \bmod p$ and sends it to Bob, who checks that this *z* is not already in his database. (If it is, that means that Alice has voted in this poll already.) Then, Alice sends Bob a complete voter-ticket (y^{r_j}, g^{r_j}) for all $j \in \overline{C}$, along with the signature *S*, which Bob verifies by finding $\prod_{j \in \overline{C}} H(y^{r_j}, g^{r_j}) \bmod n$ and checking that this is equal to $S^e \bmod n$.[3]

Finally, Alice uses the parallelized Schnorr protocol (described below) to prove to Bob the following claim: "I know α such that $y^{r_j} = (g^{r_j})^\alpha \bmod p$, and … [likewise for all $j \in \overline{C}$] … and $z = f^\alpha \bmod p$." In fact, this claim is satisfied by $\alpha = x$, but Alice does not reveal *x* in carrying out the proof.

If Bob is satisfied with the proof, he records Alice's vote and adds *z* to his database. Alice can now discard the voter-ticket (since, if she were to use it again, even in a different poll, it would be linkable to the present vote with Bob).

---

1 These parameters should be chosen such that $p = aq + 1$ for some positive integer *a*, with $q > p^{1/10}$, in order that the decisional Diffie-Hellman assumption will hold for the resulting group (Boneh, 1998). (The author would like to acknowledge Ronald Rivest for calling attention to this point.)

2 This can be found, for example, as $f = r^{(p-1)/q} \bmod p$, for some random $r \in_R \mathbb{Z}_p^*$.

3 The use of *j* and $\overline{C}$ in this section is merely a notational convention; Alice should not disclose these values to Bob, since otherwise he could almost certainly link the vote back to Alice's real identity, were he to collude with the Issuer.

4. Parallelized Schnorr ZKP protocol

Schnorr (1989) gives a protocol for constructing zero-knowledge proofs of claims of the form "I know a such that $y = g^a \bmod p$," where g is a generator of prime order in the field modulo a prime p, and y is some public value. (In general, Roman letters denote values known to both the prover and the verifier, while Greek letters denote values whose knowledge is to be zero-knowledge-proved.) Camenisch and Michels (1999) extend this protocol to conjunctions of such statements over the same α; i.e. "I know a such that $(y_1 = g_1^a) \wedge (y_2 = g_2^a) \wedge \ldots \wedge (y_k = g_k^a) \bmod p$." We will describe Camenisch and Michels's extension ("parallelized Schnorr"), and briefly analyze its security.

4.1. Description

Alice wishes to prove to Bob that she knows α, as stated above. They perform the following steps:

1. Alice chooses random $r \in_R \mathbb{Z}_p^*$, and sends $t_i = g_i^r$ to Bob, for i ∈ [1, k].
2. Bob chooses random $c \in_R [0, 2^{m-1}]$ and sends it to Alice (where m is a security parameter).
3. Alice responds with $s = r - ca$.
4. Bob accepts the proof if and only if $t_i = g_i^s y_i^c$ for all $i \in [1, k]$.

This procedure is here termed "parallelized Schnorr" because it is essentially executing Schnorr's protocol in parallel for each (y_i, g_i) pair simultaneously, using the same r, s, c for each.

Assuming that H is a good hash function, we can make this procedure non-interactive by a variant of the Fiat-Shamir heuristic (Fiat and Shamir, 1986). Instead of obtaining c from Bob, Alice calculates $c = H(t_1, t_2, \ldots t_k) \bmod 2^m$. Then, since this c is not predictable while Alice is computing her commitments, the probative effect of the interaction is the same as if Bob had chosen c randomly himself. In this way Alice can perform the entire proof in a single message.

4.2. Analysis

In analyzing this protocol, we make use of Schnorr's result, which we will not prove here:

> In the ordinary (non-parallelized) Schnorr protocol with $k = 1$, no polynomial-time adversary that does not actually know α can forge an acceptable proof with probability more than 2^{-m}, assuming that finding discrete logarithms is hard.

From this it follows that the same probability bound applies to the $k > 1$ case as well. Suppose, without loss of generality, that the best a that Alice knows is one that satisfies all of the relations but one, so $y_i = g_i^a \bmod p$ for all i except $i = 1$. Then suppose that Bob executes a "forgiving" modification of parallelized Schnorr, wherein he simply ignores the verification results $t_i \stackrel{?}{=} g_i^s y_i^c$ for i ≠ 1, and decides to accept or reject solely on the basis of $t_1 \stackrel{?}{=} g_1^s y_1^c$. This "forgiving" verification method is in fact identical to ordinary (non-parallelized) Schnorr, applied to (y_1, g_1). But, since actual parallelized Schnorr returns "reject" in strictly more cases, the probability of deceiving it must be no greater than for ordinary Schnorr.

5. Security

We now show that the protocol presented above has all three of the desired security properties (non-duplicability, anonymity, and unlinkability), under the conjunction of the discrete logarithm, decisional Diffie-Hellman, and RSA computational hardness assumptions, and assuming that the parallelized Schnorr protocol is in fact sound and zero-knowledge. (We also assume that there is more than one user, since otherwise the notions of anonymity and unlinkability are meaningless.)

5.1. Non-duplicability

Alice can double-vote if she can trick the Issuer into signing a fake voter-ticket comprised of entries $(g^{r_j}, \hat{y}^{r_j})$, where $\hat{y} = g^{\hat{x}}$ and $\hat{x} \neq x$ is some secret value known to Alice. (Assuming that H is one-way and that RSA is secure, Alice cannot outright forge these signatures.) Then, she would be able to perform the ZKP with both $z = f^x$ and $\hat{z} = f^{\hat{x}}$; and since $z \neq \hat{z}$, Bob will consider these two distinct votes.

However, Alice can only do this if, during the cut-and-choose phase of the protocol, she constructs exactly $k/2$ of the entries according to the fake form with $\hat{x}$, and these just so happen to be exactly the entries that the Issuer does not request Alice to reveal. The probability of this happening is $\binom{k}{k/2}^{-1}$, which is $O(2^{-k/2})$, small enough to ignore, especially if the Issuer rate-limits Alice after a certain number of failed challenges.[4]

Otherwise, the only way that Alice can double-vote is if she can find some $\hat{x}$ such that $y^{r_j} = (g^{r_j})^{\hat{x}} \bmod p$ for all entries in the ticket, but where $f^{\hat{x}} \neq f^x \bmod p$. This is impossible so long as g and f are both generators of the same order q.

5.2. Anonymity

The blinding scheme used in the voter-ticket issuance phase (multiplying the hashes by b_i^e, and dividing the signature by $\prod b_i$) ensures that the Issuer cannot link the issuance of a ticket with that same ticket when it is used. In other words, even if the Issuer records all interactions and colludes with Bob when Alice displays her ticket, they cannot determine which of the identities in their database the ticket corresponds to.

Each blinding factor b is chosen uniformly at random over integers relatively prime to n; and since the RSA signing/verification functions are bijective, the b^e values also have the same distribution. This means that, upon seeing $b^e h$, the Issuer learns nothing about h other than that it differs multiplicatively from $b^e h$ by some factor that is relatively prime to n; but this only narrows the possibility set by a factor of $\varphi(n)/n$, which is negligibly less than 1 (i.e. it differs from 1 by a factor that shrinks exponentially in $|n|$).

[4] Even this concern may be obviated by making k large enough; however, since modular exponentiation is rather expensive, it is preferable to keep k as small as possible.

5.3. Unlinkability

To ensure unlinkability, Alice should use a different voter-ticket every time she votes. From a single ticket it is impossible to tell what her private value x is, since this is an instance of the discrete logarithm problem. As well, given two entries (y_1, g_1) and (y_2, g_2) from two different tickets, the problem of determining whether the tickets were created using the same user secret x is at least as hard as the decisional Diffie-Hellman problem: If an algorithm A exists such that $A(y_1, g_1, y_2, g_2)$ returns "true" if and only if there exists some α such that $y_1 = g_1^{\alpha}$ and $y_2 = g_2^{\alpha} \bmod p$, then $A(g, g^a, g^b, g^{ab}) \wedge A(g, g^b, g^a, g^{ab})$ will tell us whether (g^a, g^b, g^{ab}) is a Diffie-Hellman triplet.

The same is true of the problem of determining, given two different poll bases f_1 and f_2, whether f_1^X and f_2^X were created using the same x value. An adversary can do this only if they already know the discrete logarithm of f_1 to the base of f_2 (or vice-versa); if both are chosen by a verifiable random method, then this would itself be another discrete logarithm problem.

Lastly, under the assumption that the parallelized Schnorr protocol is indeed zero-knowledge, there is no way for an adversary to tell whether two instances of the proof using separate tickets were generated using the same x value.

6. Implementation

For a reasonable security level, we set $k = 60$, $m = 80$, and the RSA key size $|n| = 2048$. We set the prime modulus and group order sizes $|p| = 2048$ and $|q| = 256$, following NIST recommendations for DSA keys (NIST, 2013). We use SHA-256 as our hash function H, on the assumption that it has the desired pseudorandom and one-way properties. We then seek to determine whether typical computer systems can execute the protocol in a reasonable amount of time.

6.1. Setup

All functions are implemented in JavaScript, using the Stanford JavaScript Crypto Library (SJCL)[5] and JavaScript Big Number (JSBN)[6] as dependencies. The code can be run in either a client (web browser) or server (NodeJS) environment.

We run two benchmark tests to time the execution of the protocol: issuing a voter-ticket (Section 3.3 above), and presenting the ticket (Section 3.4). These tests simulate the roles of all three parties one after another; there is no network communication or parallelization. The tests are repeated 20 times on a freshly-started system with no other processes running; we measure the duration (in milliseconds) of each round of testing.[7]

The tests are run on an 11-inch MacBook Air with OS X Yosemite 10.10.2, with a 1.6 GHz Intel Core i5 processor and 4 GB 1600 MHz DDR3 memory. Four different environments are used: Google Chrome 43.0.2357.65 (64-bit), Firefox 38.0.1, Tor Browser 4.5.1, and NodeJS v0.12.3.

[5] https://github.com/bitwiseshiftleft/sjcl

[6] http://www-cs-students.stanford.edu/~tjw/jsbn

[7] The test code can be accessed at https://github.com/jatchili/mcos

6.2. Results

The results are given in milliseconds:

	Issuing a ticket		Presenting a ticket	
Environment	Mean	St.Dev.	Mean	St.Dev
Chrome	3680.7	119.7	14532.8	67.6
Firefox	2962.6	53.7	11674.3	48.2
Tor Browser	2977.7	39.5	12226.4	199.1
NodeJS	3783.3	80.6	20772.7	3005.0

6.3. Analysis

These observations are pessimistic, mainly because the tests do not take advantage of parallelization. If a client and server engage in several sequential rounds of the protocol (e.g. to issue a batch of voter-tickets, or to submit votes in several different polls), then the time can be approximately halved by having the client prepare the next request while the server is verifying the last.

These results do show that implementing this protocol is well within the realm of feasibility for ordinary computer systems available today. However, in the context of web browsers, extremely liberal use of MCOS credentials – such as issuing 1,000 voter-tickets simultaneously, or using a separate ticket for every single "like" or "upvote" action on a website – is likely to be inconvenient unless one can access faster cryptographic primitives than those available in JavaScript. A more practical approach would be to use one ticket per session or user account.

7. Conclusion

The system presented here has a wide variety of immediate applications beyond straightforward polling. As mentioned above, it is useful to know that product reviews, endorsements, popularity scores, etc. are guarded against so-called "Sibyl attacks" – i.e. the use of multiple identities to create the illusion of a majority consensus.

Additionally, MCOS credentials could be used as a substitute for CAPTCHAs in preventing spam or excessive use of resources. A user could complete a single CAPTCHA presented by the credential-issuer in exchange for a batch of voter-tickets, each of which they can use to prove their humanity to another party. Then, the user will not need to complete any more CAPTCHAs until all the tickets are used up.

Lastly, MCOS credentials could solve the dilemma faced by anonymous online communities (forums, wikis, etc.), where it is effectively impossible to ban a member, since anyone can easily rejoin under a different name. If MCOS authentication is required for membership, then such behavior can be prevented without any loss of privacy.

8. References

Boneh, Dan. The Decision Diffie-Hellman Problem. Algorithmic Number Theory – Lecture Notes in Computer Science. 1998. 1423, 48-63.

Camenisch, Jan, and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. EUROCRYPT '01. 2001. 2045, 93-118.

Camenisch, Jan, and Markus Michels. Separability and Efficiency for Generic Group Signature Schemes. Advances in Cryptology – CRYPTO '99 Lecture Notes in Computer Science. 1999. 1666, 413-30.

Chaum, David. Blind Signatures for Untraceable Payments. Advances in Cryptology. 1983. 199-203.

Chaum, David, Amos Fiat, and Moni Naor. Untraceable Electronic Cash. Advances in Cryptology – CRYPTO '88 Lecture Notes in Computer Science. 1990. 403, 319-27.

Fiat, Amos, and Adi Shamir. How To Prove Yourself: Practical Solutions to Identification and Signature Problems. Advances in Cryptology – CRYPTO '86 Lecture Notes in Computer Science. 1986. 263, 186-94.

NIST. Digital Signature Standard (DSS). FIPS Publication 186-4. 2013.

Schnorr, C.P. Efficient Identification and Signatures for Smart Cards. Advances in Cryptology – CRYPTO '89 Lecture Notes in Computer Science. 1989. 435, 239-52.

Automatic Identification of Human Emotions in Facial Expressions

Henry Nassif[1], George Pantazis[1], Sebastien Boyer[2], Max Zheng Qu[3], Aude Oliva[4]

[1]Student Contributor, Class of 2016, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA 02139, USA

[2]Student Contributor, Class of 2016, MIT Engineering System Division, Cambridge, MA 02139, USA

[3]Student Contributor, PhD Candidate, MIT Department of Mechanical Engineering, Cambridge, MA 02139, USA

[4]Principal Research Scientist, Computational Perception and Cognition Group, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA 02139, USA

Intelligent Human Computer Interaction (HCI) has been the subject of increased research due to its interdisciplinary aspect and importance in shaping the boundary between man and machine. A particular field of HCI, Automatic Emotion Recognition, draws particular interest because of the complexity underlying the physical expression and manifestation of emotions, which has yet to be defined. In this paper, we design and implement an Automatic Emotion Detection System (AuDE), then analyze and compare the results obtained from different classification methods. The methods are trained and tested using cross validation on the Karolinska Directed Emotional Faces dataset (KDEF) (Lundqvist et al, 1998), and achieve an accuracy above 80%, while correctly predicting 65% of images in-the-wild. The results show that, in the scenario where the training dataset is limiting, feature-based approaches such as Support Vector Machines and Logistic Regression tend to outperform Convolutional Neural Networks.

1. Introduction

Much research has been conducted on improving and developing the interaction between humans and computers. At the core of this research one would expect automatic emotion recognition to be at the forefront because of its widespread application in affective computing, psychology, security, marketing and advertisement. To the knowledge of the authors comparatively limited work has been done in automatic emotion recognition in the field of HCI. The main objective of this research is to develop an Automatic Facial Expression Recognition System (AFERS) that is able to decode emotions hidden behind facial expressions. Our approach relies on supervised learning techniques ranging from Support Vector Machines (SVM) to Convolutional Neural Networks (CNN). We have built a system that learns from a set of labeled training images, then processes an input image and outputs a probability distribution over 7 classes of emotions: Happy, Sad, Angry, Neutral, Surprise, Disgust, Fear. The system is designed to optimize classification accuracy while minimizing computation in order to be eventually run in real time. Potential future applications would span from psychology research to human computer interaction. Specifically, this system could contribute to the humanization of robotic assistants, the study of human behavior and ubiquitous computing.

2. Related Works

Observing facial expressions is a natural way of humans to recognize emotions. Extensive studies on human expressions have laid a strong basis for the existence of universal facial expressions (Ekman, 1991). The Facial Action Coding System was introduced by Paul Ekman to cover all possible expressions in static images (Ekman, 1994). This yielded the belief that emotions could be determined systematically, independent of the individual.

In order to recognize these expressions, many feature-based algorithms have been proposed and implemented. Although different in the features they use, these algorithms follow a common hierarchy where the features are first extracted from facial images or sequence of images, and then used in a classification module.

Most algorithms rely on localization and detection of spatial features such as eyes, nose, mouth, chin, eyebrows and their geometrical relationships. Fasel and Luettin (Fasel, 2003) analyzed

facial expressions using motion extraction techniques between static images. Kumano used variable intensity templates to analyze pose-invariant facial expressions and perform classification (Kumano, 2007). In Niese et al. (2015) the authors use pose estimation and transformation of feature points coupled with a mapping model into a 2-D plane which encodes emotions based on their Valence (Positive or Negative) and their Arousal (High or Low). The plane is then discretized into 7 regions corresponding to the different emotions. In Zeng et al. (2006), Zeng treated the emotional facial expression as one class classification problem and separated it from non-emotional facial expressions.

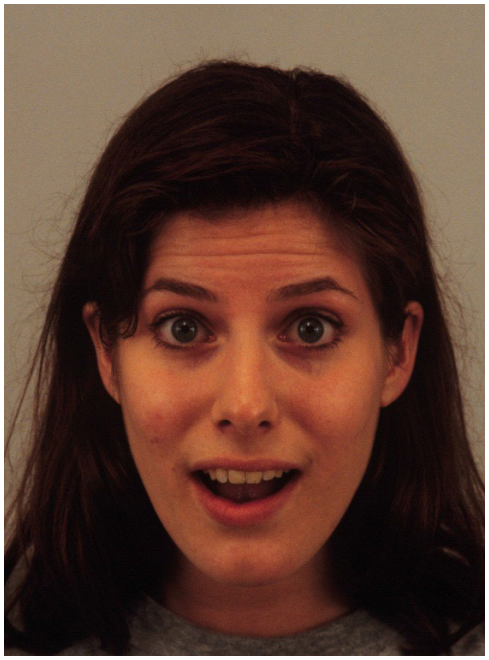


Figure 1: Sample face from the KDEF dataset showing a surprised female.

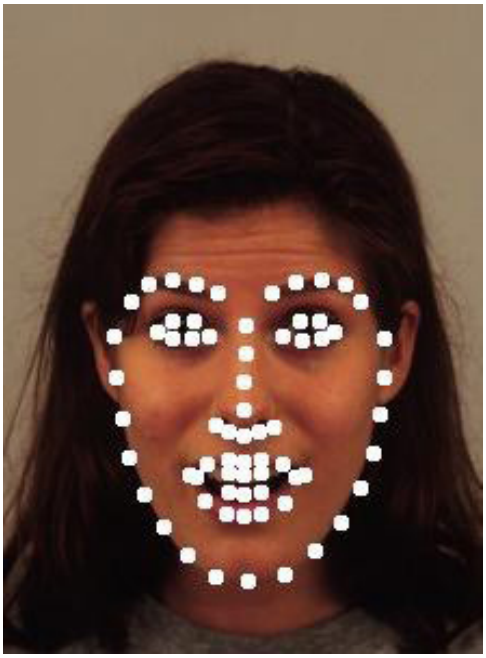


Figure 2: Face features detected by CLM-framework function.

Some other approaches are built around principal component analysis (PCA). Ahonen et al. use PCA to classify facial expressions from static images by extracting local features called Local Binary Pattern (LBP) (Ahonen, 2004). PCA-based approaches can also be combined with a Hierarchical Radial Basis Function Network (HRBFN) to classify facial expressions (Lin, 2006). In Mishra et al. (2014) the authors extend PCA to encode not only the distance to the mean face of an emotion but also the variance of the distances that encode transitions between emotions (FuzzyPCA). This approach results in a classification accuracy between 63% and 100% across emotions. In Gosavi et al. (2014), PCA is coupled with Singular Value Decomposition and a Euclidean Distance Classifier, to achieve 78.57% recognition accuracy on the Japanese Female Facial Expression (JAFFE, http://www.kasrl.org/jaffe.html) database.

As evidenced by the previous work described above, state of the art techniques have largely not been used to date for the purpose of emotion detection, while in addition demonstrating stagnant performance. Moreover, previous attempts only classified into a subset of the emotions we address, or used limited data sets to acquire results.

3. Approach

We adopt two main approaches to analyze an image dataset. The first approach relies on feature based methods, which have been primarily used in the bulk of the previous work. In addition, we combine several features to improve the accuracy of classification. The second approach relies on state-of-the-art deep learning methods, specifically Convolutional Neural Networks, which have been successful in an extensive number of other computer vision applications.

3.1 Image Dataset

In order to train and test our methods we required a dataset of facial emotions. The dataset used was "The Karolinska Directed Emotional Faces (KDEF)" dataset (Lundqvist et al, 1998). This dataset consists of 4900 pictures of human facial expressions of emotions and contains 70 individuals (35 female and 35 male), each displaying 7 different emotional expressions, with each expression photographed (twice) from 5 different angles. Each picture is a 32 bit, RGB image and was adjusted to a digital grid. The vertical and horizontal positions of eyes and mouth were adjusted to specific locations on the grid, and then cropped to a size of 562 pixels width and 762 pixels height. An example of an image from the dataset can be seen in Figure 1. We opted to only use the 980 front-facing images of our dataset. The choice of only front facing images allowed us to have access to all potential nuances that encode emotions that may have been otherwise occluded from profile pictures. All of the methods tested, and presented in the following sections use the same, front-facing subset of the full dataset.

3.2 Preprocessing of Data

In order to prepare our dataset to be used on our various methods, we preprocessed the data in three different ways. The first step was to extract the facial features using the CLM-framework (Baltrusaitis, 2013). For each still face image, the feature detection function can recognize the edges of the face contour, the eyes, nose, and mouth, as shown in Figure 2.

The feature detection method returns a 68×2 matrix with the coordinates of the detected feature points. The order in which these feature points are returned is the same, and can be

mapped to specific regions on the face (face outline, eyes, nose, mouth) based on their index. These feature points are used in the various methods we later explore. In addition they are used for additional preprocessing, which we detail in subsequent sections.

The next preprocessing step is to crop out the face from the rest of the image. These cropped images are saved separately and are also used in our methods in conjunction with the original full dataset. This cropping is achieved by using the feature points extracted using the CLM framework to form the boundary of the cropping region.

3.3 Feature-Based Methods

The feature-based approach relies on using features extracted from the image to train a low-dimensional supervised learning algorithm. This algorithm is then used to classify the emotions. We describe our approach at two stages: first at the feature extraction stage, and second at the classification stage.

3.3.1 Feature Extraction

Our approach explored the use of three main features:

- Coordinates of the features points (extracted as described in the pre-processing step)
- HOG features of the cropped image
- Principal Components of the cropped image.

We tuned the parameters for each of the features based on the 10-fold cross validation error. The best choice for the number of dimensions of each of the features is shown in Table 1.

Feature name	Min	Max	Best
Points coordinates	136	136	136
HOG features	~50	~5000	324
PCA features	1	980	60

Table 1: Range and best dimensions for the 3 types of features used in the extraction process.

Figure 3 and 4 display the results of the logistic regression using different parameters. As can be seen, Principal Component Analysis (PCA) yields the best result when using 60 principal components. Conversely, a HOG window of size 15 pixels (corresponding to a total of 324 features) yields the best result.

3.3.2 Classification algorithm

Once the features have been extracted, we benchmarked three common classification algorithms:

- Logistic regression
- SVM
- Random forest

The duration of training varied substantially from one classifier to another. SVM and Logistic regression could be trained quickly, but random forest often took several minutes. Similar to our method for the choice of the hyper parameters of

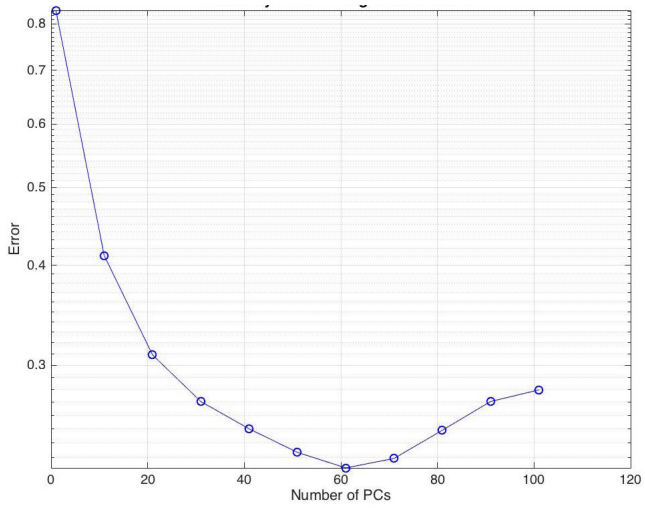


Figure 3: Tuning the number of principal components (PCs).

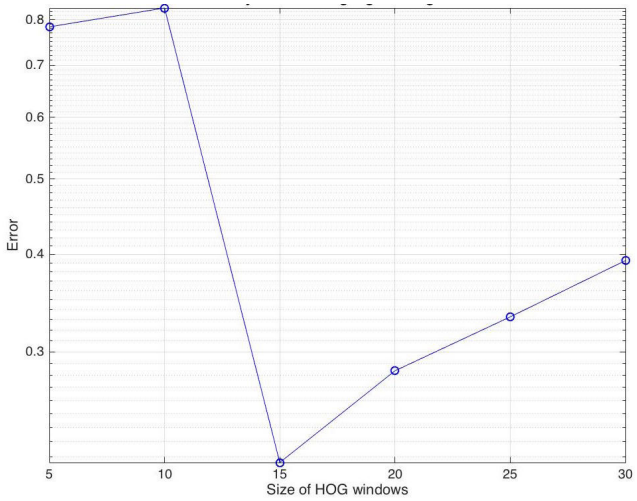


Figure 4: Tuning the size of the HOG box.

the feature extraction process, we used 10-fold cross-validation to choose the best algorithms. Figure 5 shows the error for different values of the number of trees in the random forest algorithm when training on 100 Principal Components. The graph shows that the lowest error is obtained when using either 350 or 550 trees.

3.3.3 Comparison of Results

After optimizing all the hyper-parameters individually, we focused on benchmarking all the possible combinations of features and classifiers. Here we used fixed-size images of dimensions 64x64. These results are reported in Table 2 and show that our best algorithm is mixing HOG features with either the point coordinates or the PCA features. This combination achieves an error rate of 17% (average over 10-fold cross validation) when trained using a random forest algorithm.

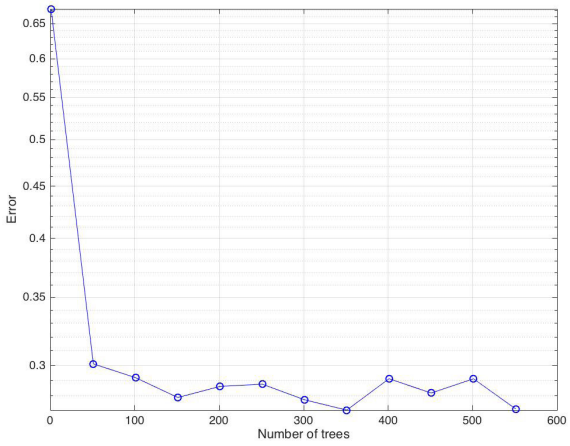


Figure 5: Average error 10 fold cross validation using 100 PC.

	SVM	Log-Reg	R-Forest
Points	36%	31%	31%
PCA	33%	23%	32%
HOG	25%	23%	24%
Points + PCA	33%	21%	33%
HOG + PCA	22%	22%	**17%**
HOG + Points	21%	24%	**17%**
Points + PCA + HOG	20%	21%	18%

Table 2: Average error after the 10-fold cross validation for input images of size 64x64. R-Forest refers to the Random forest algorithms (used with 200 trees).

3.4 Convolutional Neural Networks Method

In the subsequent part, we investigated different CNN structures with the goal of achieving better classification results. The CNNs were trained using cross-validation, specifically training on 80% of the data and testing on the remainder 20%.

In the first experiment we compared the performance of the CIFAR network on a 3-channel cropped image vs 4-channel cropped image, where the original image was augmented with the coordinates of the feature points represented as a binary matrix in a 4th Channel (3 RGB+1 Feature points). Table 3 below shows the structure of the CIFAR network – a 5 layer

Layer	Structure
1st layer	32 5x5x3 + pool 3x3 + relu
2nd layer	32 5x5x32 + relu + pool 3x3
3rd layer	64 5x5x32 + relu + pool 3x3
4th layer	64 4x4x64 + relu
5th layer	7 1x1x64 + softmax

Table 3: CIFAR structure.

CNN. Figure 6 and 7 show the training and validation errors for CIFAR-3D and CIFAR-4D respectively, where we observe a top 1 validation error of 20% and a top 3 validation error of less than 5% for both approaches.

Although the training error is smaller for the 4D channel, the results show that the CNN performance is comparable on the 3D inputs and the 4D inputs. This might be due to the fact that the feature content of the 4th channel is not optimal, and potentially including a different set of features might result in a better classification score.

Subsequently, we designed a different network with an architecture similar to CIFAR, but with a denser layers. Table 4 below shows the structure of the Denser CIFAR network. Figure 8 shows that this network did indeed perform slightly better than the previous CIFAR with top 1 validation error just shy of 20% and top 3 validation error of approximately 2%.

4. Experimental Results

We first evaluate the performance of the different algorithms using the average error on a 10-fold cross validation run. We then used a different set of images (derived from the web or self-generated) to assess our algorithm's ability to cope with new data (testing in-the-wild).

4.1 Baseline

The lowest reasonable baseline is the average performance of a "random guess". Since our dataset contains 7 classes of equal size (140 images per class), the expected error rate is 86%. In practice, we expect our classification to have a much smaller error rate than 86%

4.2 Error of "best guess" on 10-fold cross validation

Table 5 displays the performance of our two approaches. We observe that for images of size 32x32, the CNN method performs better than our feature-based approach. However, the relative difference is smaller than the difference usually observed in similar computer vision tasks. We think this could mainly be explained by the restricted size of our data set. As the size of the input image increases, the performance of the CNN decreases while that of the feature-based method increases.

At the time of publication, our best model is a feature-based model, used on images of bigger size (128x128). Our hypothesis is that the restricted size of the dataset limits the learning of the CNN to smaller number of features (smaller images). Indeed, when the size of the images increase, the learning of the weights becomes more complex, requiring more training samples. In addition, using a small dataset becomes more and more problematic when the number of weights increases.

We then analyzed the accuracy of our algorithm for each of the 7 classes and concluded that whereas certain classes are easily identified (94% accuracy for the "Happy" class), other classes were often harder, resulting in confusion (such as "Afraid" being often misclassified as "Surprised"). Figure 9 displays the confusion matrix for the seven classes, and

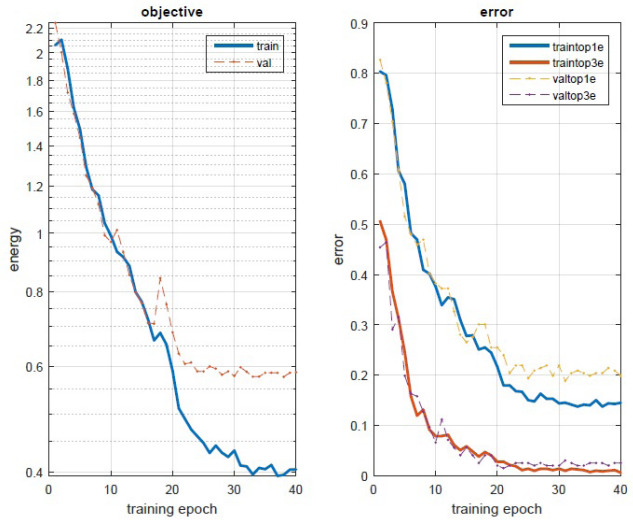


Figure 6: Training and Validation error for CIFAR structure on 3D image of size 32x32.

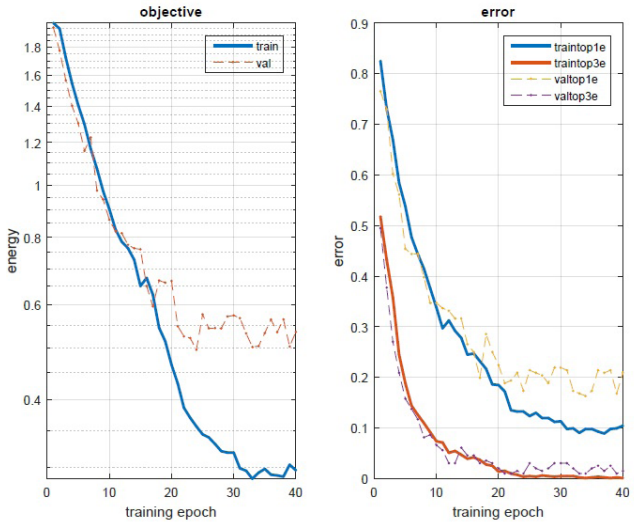


Figure 7: Training and Validation for CIFAR on 4D image of size 32x32.

Layer	Structure
1st layer	64 5x5x3 + pool 3x3 + relu
2nd layer	64 5x5x64 + relu + pool 3x3
3rd layer	64 5x5x64 + relu + pool 3x3
4th layer	64 4x4x64 + relu
5th layer	7 1x1x64 + softmax

Table 4: Structure of dense CIFAR.

Image size	FBM	CNN
32x32	25%	20%
64x64	17%	25%
128x128	15%	N/A

Table 5: Average error after the 10-fold cross validation for the best hyper-parameters (FBM: Feature-Based Methods)

shows that high arousal emotions (i.e. highly expressive) such as "Happy", "Disgusted" and "Surprised" seem to be easier to classify correctly. The "Neutral" emotion appears to also be easily identifiable perhaps due its clear distinction from other emotions.

4.3 Performance on images outside the Dataset

The next step of the evaluation is to benchmark our algorithm's performance on images outside the data set used. We sample images from the web and generate some independently. We then perform the preprocessing steps explained earlier and input the preprocessed images into our algorithm for testing. We used human judgement to decide the "true" labels of images. We summarize the transferability of our algorithm in Table 6.

	Initial Dataset	Web
Random Guess	86%	86%
CNN	17%	35%

Table 6: Error of our best algorithm (trained on the Initial Dataset) on both the Initial dataset and a dataset of images from the web. The error for the initial dataset is the average of the 10-fold cross validation error (not the error on the train data Per Se).

On 20 images, our algorithm identified 13 correctly (the best guess was the same as the human estimation), leading to an error rate of 35%. Figure 10 displays examples of the images use for this final step of testing.

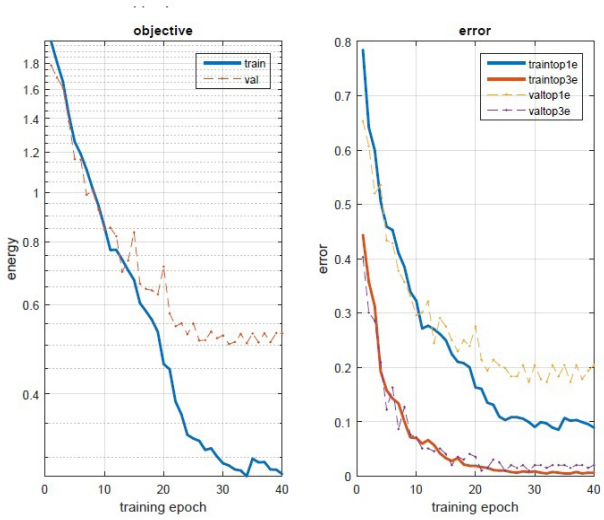


Figure 8: Training and Validation for Dense CIFAR on 3D image.

5. Conclusion

This paper provides an overview of automatic emotion recognition using different state-of-the-art techniques. The techniques are individually optimized and then combined to produce the best classification results on the Karolinska Directed Emotional Faces dataset (KDEF) (Lundqvist et al, 1998). The results show that there is no single technique that

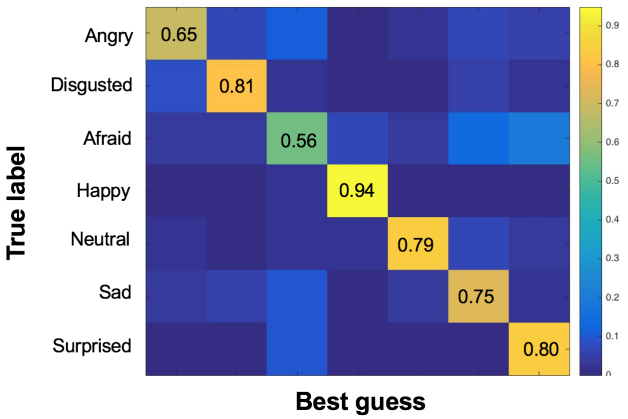


Figure 9: Confusion matrix of the best predictive algorithm.

always performs better than others, but that the performance of these methods is very much dependent on the characteristics of input data, such as size of dataset, size of image and image content. The main limitation of the system is the small size of the training dataset which prevented the CNN from achieving full potential. On small images the CNN performed better than feature-based methods, but on larger images feature-based methods outperformed the CNN. The system can be further improved by exploring different face segmentation methods (potentially running CNN on subfaces rather than an entire face) and extending it to work on profile views, not just front-facing views. In addition the CNN performance can be improved by augmenting the input images with different feature channels as well as optimizing computations during the training phase. Furthermore, this work can be extended to identify emotions based on a continuum rather than just discrete categories. Lastly, we would like to extend our approach to work on video data so as to be applicable in real time.

6. Acknowledgements

We would like to thank Professor Aude Oliva and Yusuf Aytar for providing us with the knowledge to complete this research.

We would also like to thank Aditya Khosla, Bolei Zhou and Harini Kannan for answering our questions, and pointing us to the right resources we needed.

7. References

[1] P. Ekman and M. O'Sullivan, Facial expression: methods, means, and moues. Cambridge University Press, New York, 1991.

[2] R. Avent, C. Ng, and J. Neal, "Machine vision recognition of facial affect using backpropagation neural networks," in Pro-

[3] R. Niese and et al, "Machine vision based recognition of emotions using the circumplex model of affect," in 2011 International Conference on Multimedia Technology, p. n, 2015.

[4] S. R. Mishra, B. Ravikiran, K. S. M. Sudhan, N. Anudeep, and G. Jagdish, "Human emotion classification using fuzzy and pca approach," in Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013, pp. 75–81, 2014.

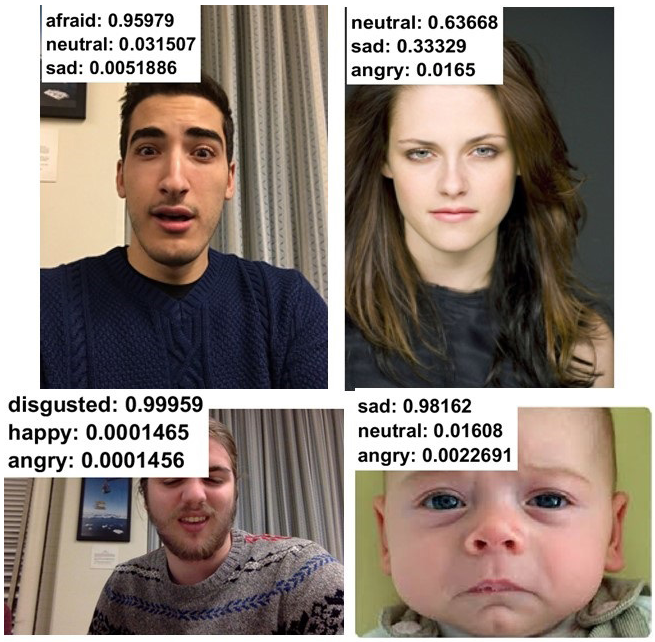


Figure 10: Examples of images and the top three corresponding predictions.

[5] A. Gosavi and S. Khot, "Emotion recognition using principal component analysis with singular value decomposition," in Electronics and Communication Systems (ICECS), 2014 International Conference on, pp. 1–5, Feb 2014.

[6] D. Lundqvist, A. Flykt, and A. Ohman, The Karolinska Directed Emotional Faces - KDEF. CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998, http://www.emotionlab.se/resources/kdef.

[7] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, pp. 354–361, Dec 2013.

[8] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," Pattern Recognition, vol. 36, no. 1, pp. 259-275, 2003

[9] Ekman, P.: Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychol. Bull. 115(2), 268–287, 1994.

[10] D. Lin, "Facial expression classification using PCA and hierarchical radial basis function network," Journal of information science and engineering, vol. 22, no. 5, pp. 1033-1046, 2006.

[11] T. a. H. A. a. P. M. Ahonen, "Face recognition with local binary patterns," in European Conference on Computer Vision, 2004.

[12] Kumano, S., Otsuka, K., Yamato, J., Eisaku, S., Sata, Y.: Pose-Invariant facial expression recognition using variable intensity templates. Asian Conf. on Computer Vision, 2007.

[13] Zeng, Z., Fu, Y., Roisman, G.I., Zhen, W.: Spontaneous emotional facial expression detection. Journal of Multimedia, 2006.

[14] Bartlett, S., Littlewort, G., Frank, G., Lainscsek, C., Fasel, I., Movellan, J.: Fully automatic facial action recognition in spontaneous behavior. In Proc. Conf. Automatic Face & Gesture Recognition, pp. 223--230, 2006.

One focus: a shared commitment to improve the lives of cancer patients everywhere.

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.**

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

@TakedaOncology

Takeda ONCOLOGY