



How to Make Sense of Data

Caltech scientists and engineers are designing the sensors that gather data and the algorithms that translate data to communicate and work together in a process they call Sensing to Intelligence (S2I).

By Andrew Moseman



In April 2017, a group of observatories scattered across the face of the earth, from Spain to Hawaii, turned their attention to one object: a black hole called M87 that lies around 55 million light-years away. The radio instruments that make up this global network of observatories, known as the Event Horizon Telescope (EHT), gathered data about M87 throughout an entire night of observation; EHT researchers then fed that mountain of information into algorithms that processed it, adjusted for the changing sky and the distances between telescopes, and finally created the historic first image of a black hole.

Kathryn L. Bouman, who worked on that black hole image before joining Caltech as assistant professor of computing and mathematical sciences, electrical engineering, and astronomy, now collaborates with a team to create the Next-Generation Event Horizon Telescope (ngEHT) that will bring more black holes into view and increase the understanding of these bizarre cosmic phenomena. To realize such an instrument, which will include Caltech's Owens Valley Radio Observatory, the scientists plan to reverse their typical process: rather than simply building a bigger network of observatories and asking a new generation of algorithms to make sense of the data, Bouman and colleagues plan to use their algorithms to find weak spots in the current data and find the optimal places to build new telescopes and thereby enhance their ability to see black holes. The machine-learning side of EHT is not only interpreting data, it is informing the design and operation of the sensors themselves. "The goal is to try to combine sensing and intelligence," Bouman says, "to no longer separate them, but to think of this as a whole pipeline."

Sensors and machine learning lie at the forefront of research across the sciences. At any given moment, billions of sensors around the world, ranging from something as simple as a motion detector for a home to something as complex as the LIGO (Laser Interferometer Gravitational-wave Observatory) detectors, generate a pile of information whose size increases exponentially. Sensors allow seismologists, climate change experts, and gravitational-wave hunters to amass enormous data sets, while computer algorithms and human researchers work unceasingly to sort the signal (the true information) from the noise in this data stream. These two fields are typically seen as distinct, but the ngEHT is one example of how sensing (the hardware) and intelligence (the software) can converge to create networks of sensors that are intelligent in their own right.

This convergence powers a new research thrust at Caltech: Sensing to Intelligence, or S2I, an initiative that unites researchers from across interests and disciplines



with a focus on co-designing sensors and learning algorithms. The researchers are, in other words, building the two sides to communicate and collaborate from the outset as they envision new ways to refine earthquake prediction, visualize distant cosmic phenomena, and find signals about a person's health in the molecules found in their breath.

"We have a unique opportunity at Caltech to transform the way we connect the information world with the physical world," says Azita Emami, the Andrew and Peggy Cherng Professor of Electrical Engineering and Medical Engineering and one of the leaders of the S2I initiative. "That can lead to a whole new generation of smart devices and instruments."

Breath Biomarkers

"Think about breath analysis as another version of a blood test," says Alireza Marandi, assistant professor of electrical engineering and applied physics. A person's breath abounds with information, with biomarkers that have the potential to inform doctors about different aspects of that patient's health. But to isolate and identify the various molecules in a cloud of vapor requires sophisticated technology that would be problematic to downsize into a device compact enough for people to use



Left: Kathryn L. Bouman
Right: In 2019, the Event Horizon Telescope (EHT) captured the first-ever picture of a black hole and its shadow by imaging the one at the center of M87, a galaxy 55 million light-years from Earth. The EHT's successor could reveal the black hole at the heart of the Milky Way.

at home. For example, Marandi says, picture a breath analyzer small enough to attach to a smartphone so that users could easily track health data over time simply by exhaling into the device. He seeks to make such technology possible by advancing the sensing and intelligence sides of the problem simultaneously.

Marandi's lab investigates how nonlinear photonics, a field of optics in which light (usually a laser) emerges with different frequencies and properties after passing through a material, could be used in sensing applications to accomplish tasks no other technologies can accomplish. For example, Marandi's team has found that laser pulses in the mid-infrared range have just the right resonant frequency to cause molecules they touch to "jiggle." Those excited molecules emit information that acts as a fingerprint, which theoretically would allow researchers to identify a particular molecule's presence in a cloud of gas, like an exhaled breath from a person's lungs. The sensing challenge, however, is that most lasers work in the visible light or near-infrared parts of the spectrum, and so most of the backbone technologies needed for that kind of a breath sensor are focused only on those ranges, too. To make use of these affordable and available tools, Marandi starts with laser light in the visible or near-infrared range and transforms his beam into the mid-infrared range in which it is most effective.

The most straightforward way to find out what molecules are inside a person's breath could be spectroscopy, a long-standing technique in which researchers look at which frequencies of light are absorbed as the beam passes through matter. However, building a perfect spectrometer would add weight and complexity to the device, and it is not necessary in this case since Marandi does not need to identify every molecule in a person's breath. Instead, he is working to develop a sensor that works only in the mid-infrared range and can ID certain noteworthy molecules that a user might need to track. The machine-learning element would extract only the relevant data and leave behind the extraneous information. To Marandi, his solution is an example of when the perfect is the enemy of the good. "You actually want to think about not having a very clean set of information but being able to get bits and pieces," he says.

Such a device would not only be more accurate than a sensor that tracks every molecule in a puff of air, but it would also be smaller and more efficient, allowing it potentially to be scaled down for consumer use. "I don't have to spend a lot of time making a full-fledged spectrometer," Marandi says. "I just need to make something from which I can extract the right information. That is one of the themes of S2I."

Join Marandi for a lecture and discussion on his work at the November 11 Watson Lecture. events.caltech.edu

The Body and the Brain

A microscopic sensor, a single tiny chip, floats through the body, protected by a thin but porous layer of platinum. Glucose and oxygen molecules can squeeze through the holes in this layer, but the body's immune cells, which would seek to destroy such a synthetic intruder, cannot pass. Once past the platinum, glucose reacts with an electrode on the surface of the chip to generate a small electrical current. This tiny chip amplifies that current and digitizes it into data representing the level of glucose in the bloodstream. The sensor then embeds these data within a radio signal to send to a receiving device outside the body, thereby providing a measurement of a person's blood sugar.



This glucose sensor, which takes the place of the frequent needle pricks diabetic patients typically must endure, was built by Azita Emami and Axel Scherer, Caltech's Bernard Neches Professor of Electrical Engineering, Applied Physics and Physics. Their device, which has been tested in mice, demonstrates the ability of low-power, implantable, tiny medical devices to become flexible technology platforms that can be adapted to a wide variety of medical uses. At the same time, algorithms can help those devices adapt to each individual's unique biomarkers and baseline health data. Feedback from this intelligence side of the system may tell the sensors to change what they measure inside the body or how they measure it.

"Imagine a two-way flow of information, not just from the sensors to the software but from the algorithm side back to the sensor," says Emami, who also serves as an investigator in the Heritage Medical Research Institute. "If you design these two together, you can create systems that are far more efficient and have better performance. Perhaps then we will be able to see things that could not be seen before."

Emami is also trying to apply S2I to perhaps the most complicated part of the body in which to put an implantable device: the brain. Caltech researchers and others have built brain-machine interface devices that can read the neural activity that occurs when a person tries to move a limb, for instance. In this way, patients with spinal-cord injuries can learn to move prosthetic devices with nothing but their thoughts. The problem, however, is that the brain is a busy environment, and it is difficult to isolate a person's intention to move a limb in a certain direction from among the myriad other neural signals. "We have huge amounts of data recorded by small electrodes in the brain, and they vary over time because of small movements by the electrodes," Emami says. "To address that, we are using algorithms that provide a more robust prediction of the intention of the patient."

Frontiers Large and Small

M87, the subject of the momentous black hole image by Bouman and colleagues, is a behemoth. It contains 6.5 billion times as much mass as the sun. Because the black hole is so vast, the light orbiting M87 requires longer to complete an orbit than it would to circle a smaller black hole. The upshot is that M87's appearance in the night sky changes

Azita Emami holds a large-scale model of her implantable health monitor. Such a device could track a person's blood glucose level without the need for daily needle pricks and could talk to machine-learning algorithms that help the sensor capture a more complete picture of a patient's health.

relatively slowly. If it went too much faster, the EHT's sensors and algorithms, already pushed to their limits, would not have been able to capture a true image of the black hole, says Bouman, one of Caltech's Rosenberg Scholars. The next-generation Event Horizon Telescope, designed through an S2I approach in which algorithms optimize the locations of new observatories, could have enough data-gathering and processing power to image another black hole: the one at the center of the Milky Way galaxy, which is far smaller and evolves much faster than M87.

Black hole research is only one example of computational imaging, a cross-disciplinary approach that creates images from data sets rather than by gathering visual light through a lens or other filter. Its practical applications are already familiar to many: ultrasounds, CT scans, and other forms of medical imagery create pictures by processing data in various wavelengths. But computational imaging can visualize subjects from the astronomical to the microscopic. For example, one of Bouman's postdoctoral researchers, Aviad Levis, looks inside clouds via tomography, the same kind of X-ray technology at work in CT scans. "Clouds may be the most important part of climate models," Bouman says. "If you want to have accurate climate models, you have to understand clouds and their microphysics." Because of clouds' complexity, researchers like Levis rely on machine learning to help them process data taken from multiple angles into a realistic model of a cloud.

Another of Bouman's students, Angela Gao, collaborates with Zachary Ross, assistant professor of geophysics, and Yisong Yue, professor of computing and mathematical sciences, on machine learning to help build a better picture of earthquakes. Often, earthquake analyses build upon a simulated model of the ground (its structure, materials, and faults) that has been computed from earlier quakes but may not be accurate for smaller quakes, Bouman says. Gao seeks to use data from a quake to build a new, better model.

Computational imaging also can lead scientists to new advances at smaller scales. Changhui Yang, Caltech's Thomas G. Myers Professor of Electrical Engineering, Bioengineering, and Medical Engineering, applies the power of machine learning to microscopy and medical imaging. For example, when pathologists look through a microscope they are trained to look for particular cells, such as cancer cells. However, an image contains far more

information; the problem is how to know what to look for. That is where machine learning enters the picture. Yang feeds microscope images to a neural network, a kind of artificial intelligence based loosely on the brain. Because he does not label the parts of the picture or tell the AI what to look for, the system is forced to figure out how to

"Imagine a two-way flow of information, not just from the sensors to the software but from the algorithm side back to the sensor."

make sense of the images and find patterns within them without the cognitive bias of a human scientist who has been trained on what to look for. "The idea here is that it's possible to build a system that can optimally collect information even if the information is not collected in such a way that a human eye can readily interpret it," Yang says.

A major challenge of machine learning, however, is that while the algorithms can process far more data than a person can, and might find patterns a human would not notice, AI cannot explain how it "thinks" or how it interprets information; and of course it is important for researchers to learn why the algorithm focused on a particular element or whether that element is important to a person's health.

Here again, Caltech researchers see the opportunity for the algorithms to inform the sensors. In the case of medical microscopy, Yang might have the neural network study certain locations within an image and then tweak the sensors (in this case, the microscopes) to match. "It's like talking to somebody who might have difficulty communicating," Yang says. "You can pay attention to where that person's eyes are focused, and that will give you some glimpse into what is going on in the person's thought process. Similarly, we can optimize the collection system to suit what the machine learning is focusing on."

Across Caltech, interdisciplinary projects of this nature that fall under the S2I framework will begin to help humanity make sense of the modern-day data deluge. "We have these two isolated domains, the sensors and the algorithms, but we believe that this is not sustainable," Emami says. "We have dealt with the exponential growth in the data from different sensors in a very blind way. It will be completely different if we co-design sensing and algorithms so that it is all done in a more intelligent way." 📍