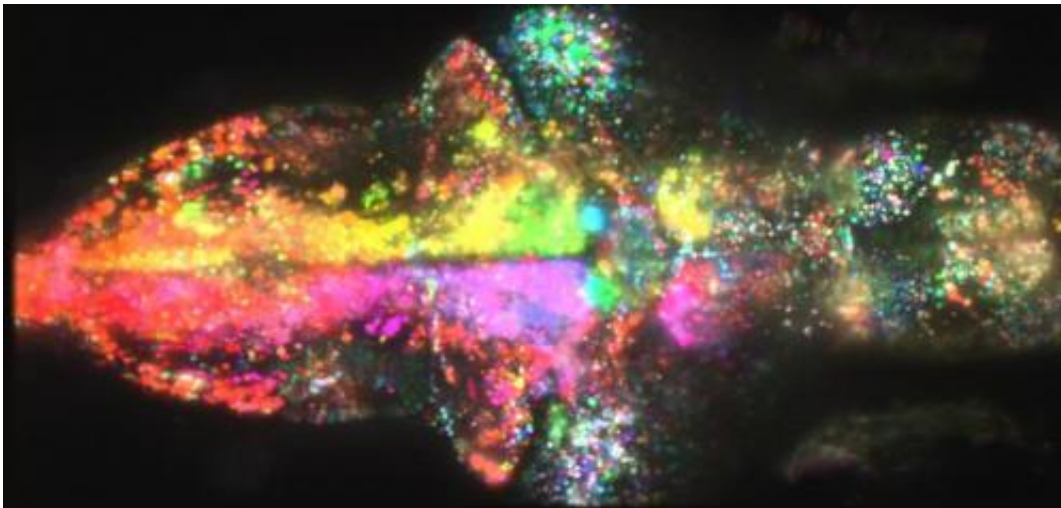


New tools help neuroscientists analyze 'big data'

July 27 2014



Whole-brain map of direction selectivity in larval zebrafish. Colors indicate the relationship between neuronal responses and the direction of a moving stimulus (yellow left, purple right, red forwards). Credit: Jeremy Freeman, Nikita Vladimirov, Takashi Kawashima, Yu Mu, Nicholas Sofroniew, Davis Bennett, Joshua Rosen, Chao-Tsung Yang, Loren Looger, Philipp Keller, Misha Ahrens

In an age of "big data," a single computer cannot always find the solution a user wants. Computational tasks must instead be distributed across a cluster of computers that analyze a massive data set together. It's how Facebook and Google mine your web history to present you with targeted ads, and how Amazon and Netflix recommend your next favorite book or movie. But big data is about more than just marketing.

New technologies for monitoring [brain](#) activity are generating unprecedented quantities of information. That data may hold new insights into how the brain works – but only if researchers can interpret it. To help make sense of the data, neuroscientists can now harness the power of distributed computing with Thunder, a library of tools developed at the Howard Hughes Medical Institute's Janelia Research Campus.

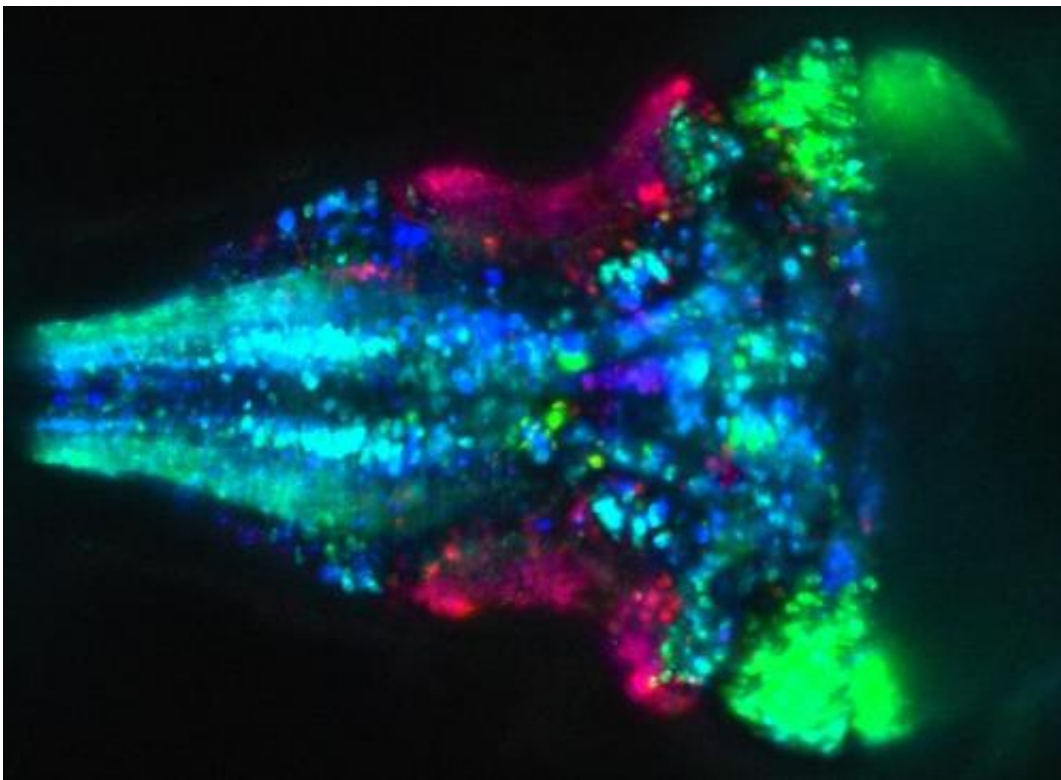
Thunder speeds the analysis of data sets that are so large and complex they would take days or weeks to analyze on a single workstation – if a single workstation could do it at all. Janelia group leaders Jeremy Freeman, Misha Ahrens, and other colleagues at Janelia and the University of California, Berkeley, report in the July 27, 2014, issue of the journal *Nature Methods* that they have used Thunder to quickly find patterns in high-resolution images collected from the brains of active zebrafish and mice with multiple imaging techniques.

Importantly, they have used Thunder to analyze imaging data from a new microscope that Ahrens and colleagues developed to monitor the activity of nearly every individual cell in the brain of a zebrafish as it behaves in response to visual stimuli. That technology is described in a companion paper published in the same issue of *Nature Methods*.

Thunder can run on a private cluster or on Amazon's cloud computing services. Researchers can find everything they need to begin using the open source library of tools at <http://freeman-lab.github.io/thunder>

New microscopes are capturing images of the brain faster, with better spatial resolution, and across wider regions of the brain than ever before. Yet all that detail comes encrypted in gigabytes or even terabytes of data. On a single workstation, simple calculations can take hours. "For a lot of these data sets, a single machine is just not going to cut it," Freeman says.

It's not just the sheer volume of data that exceeds the limits of a single computer, Freeman and Ahrens say, but also its complexity. "When you record information from the brain, you don't know the best way to get the information that you need out of it. Every data set is different. You have ideas, but whether or not they generate insights is an open question until you actually apply them," says Ahrens.



Whole-brain map of response dynamics during a stimulus-driven behavior. The analysis uses large-scale spatio-temporal factorization to reveal the timing of responses relative to the onset of a stimulus. Credit: Jeremy Freeman, Nikita Vladimirov, Takashi Kawashima, Yu Mu, Nicholas Sofroniew, Davis Bennett, Joshua Rosen, Chao-Tsung Yang, Loren Looger, Philipp Keller, Misha Ahrens

Neuroscientists rarely arrive at new insights about the brain the first time they consider their data, he explains. Instead, an initial analysis may hint

at a more promising approach, and with a few adjustments and a new computational analysis, the data may begin to look more meaningful. "Being able to apply these analyses quickly—one after the other—is important. Speed gives a researcher more flexibility to explore and get new ideas."

That's why trying to analyze neuroscience data with slow computational tools can be so frustrating. "For some analyses, you can load the data, start it running, and then come back the next day," Freeman says. "But if you need to tweak the analysis and run it again, then you have to wait another night." For larger data sets, the lag time might be weeks or months.

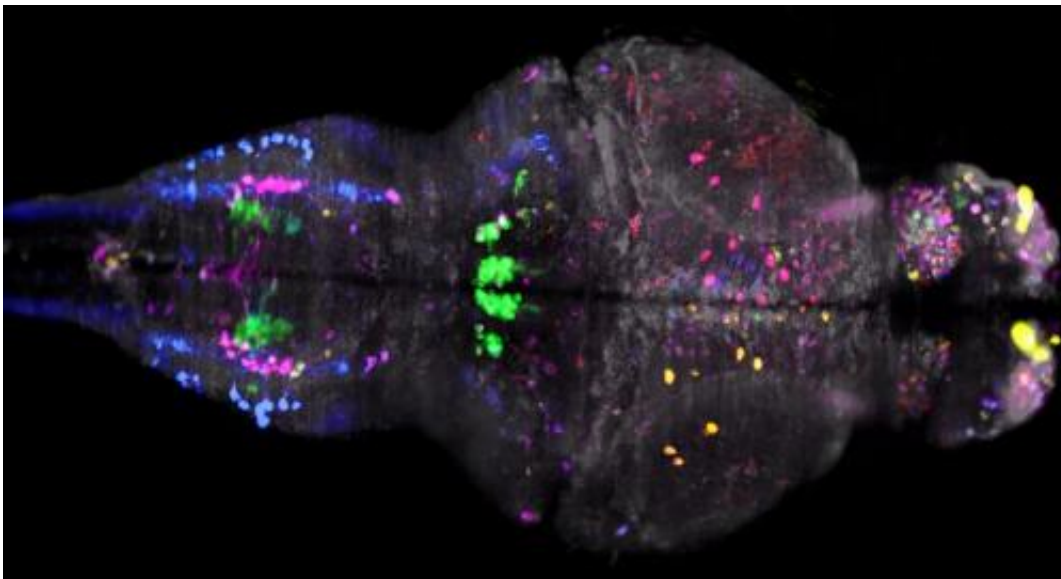
Distributed computing was an obvious solution to accelerate analysis while exploring the full richness of a data set, but many alternatives are available. Freeman chose to build on a new platform called Spark. Developed at the University of California, Berkeley's AMPLab, Spark is rapidly becoming a favored tool for large-scale computing across industry, Freeman says. Spark's capabilities for data caching eliminates the bottleneck of loading a complete data set for all but the initial step, making it well-suited for interactive, exploratory analysis, and for complex algorithms requiring repeated operations on the same data. And Spark's elegant and versatile application programming interfaces (APIs) help simplify development. Thunder uses the Python API, which Freeman hopes will make it particularly easy for others to adopt, given Python's increasing use in neuroscience and data science.

To make Spark suitable for analyzing a broad range of neuroscience data – information about connectivity and activity collected from different organisms and with different techniques – Freeman first developed standardized representations of data that were amenable to distributed computing. He then worked to express typical neuroscience workflows into the computational language of Spark.

From there, he says, the biological questions that he and his colleagues were curious about drove development. "We started with our questions about the biology, then came up with the analyses and developed the tools," he says.

The result is a modular set of tools that will expand as the Janelia team—and the neuroscience community—add new components. "The analyses we developed are building blocks," says Ahrens. "The development of new analyses for interpreting large-scale recording is an active field and goes hand-in-hand with the development of resources for large-scale computing and imaging. The algorithms in our paper are a starting point."

Using Thunder, Freeman, Ahrens, and their colleagues analyzed images of the brain in minutes, interacting with and revising analyses without the lengthy delays associated with previous methods. In images taken of a mouse brain with a two-photon microscope, for example, the team found cells in the brain whose activity varied with running speed.



Functional networks discovered using a large-scale unsupervised learning

algorithm. Different colors are different neuronal populations, superimposed on an anatomical image. Credit: Jeremy Freeman, Nikita Vladimirov, Takashi Kawashima, Yu Mu, Nicholas Sofroniew, Davis Bennett, Joshua Rosen, Chao-Tsung Yang, Loren Looger, Philipp Keller, Misha Ahrens

For analyzing much larger data sets, tools such as Thunder are not just helpful, they are essential, the scientists say. This is true for the information collected by the new microscope that Ahrens and colleagues developed for monitoring whole-brain activity in response to visual stimuli.

Last year, Ahrens and Janelia group leader Phillip Keller used high-speed light-sheet imaging to engineer a microscope that captures neuronal activity cell by cell across nearly the entire brain of an immature zebrafish. That microscope produced stunning images of neurons in the zebrafish brain firing while the fish was inactive. But Ahrens wanted to use the technology to study the brain's activity in more complex situations. Now, the team has combined their original technology with a virtual-reality swim simulator that Ahrens previously developed to provide fish with visual feedback that simulates movement.

In a light sheet microscope, a sheet of laser light scans across a sample, illuminating a thin section at a time. To enable a fish in the microscope to see and respond to its virtual-reality environment, Ahrens' team needed to protect its eyes. So they programmed the laser to quickly shut off when its light sheet approaches the eye and restart once the area is cleared. Then they introduced a second laser that scans the sample from a different angle to ensure that the region of the brain behind the eyes is imaged. Together, the two lasers image the brain with nearly complete coverage without interfering with the animal's vision.

Combining these two technologies lets Ahrens monitor activity throughout the brain as a fish adjusts its behavior based on the sensory information it receives. The technique generates about a terabyte of data in an hour – presenting a data analysis challenge that helped motivate the development of Thunder. When Freeman and Ahrens applied their new tools to the data, patterns quickly emerged. As examples, they identified cells whose activity was associated with movement in particular directions and cells that fired specifically when the fish was at rest, and were able to characterize the dynamics of those cells' activities. Example analyses like these, and example data sets, are available at the website <http://research.janelia.org/zebrafish/>.

Ahrens now plans to explore more complex questions using the new technology, and both he and Freeman foresee expansion of Thunder. "At every level, this is really just the beginning," Freeman says.

More information: *Nature Methods* [DOI: 10.1038/nmeth.3041](https://doi.org/10.1038/nmeth.3041) , [DOI: 10.1038/nmeth.3040](https://doi.org/10.1038/nmeth.3040)

Provided by Howard Hughes Medical Institute

Citation: New tools help neuroscientists analyze 'big data' (2014, July 27) retrieved 5 May 2024 from <https://medicalxpress.com/news/2014-07-tools-neuroscientists-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
