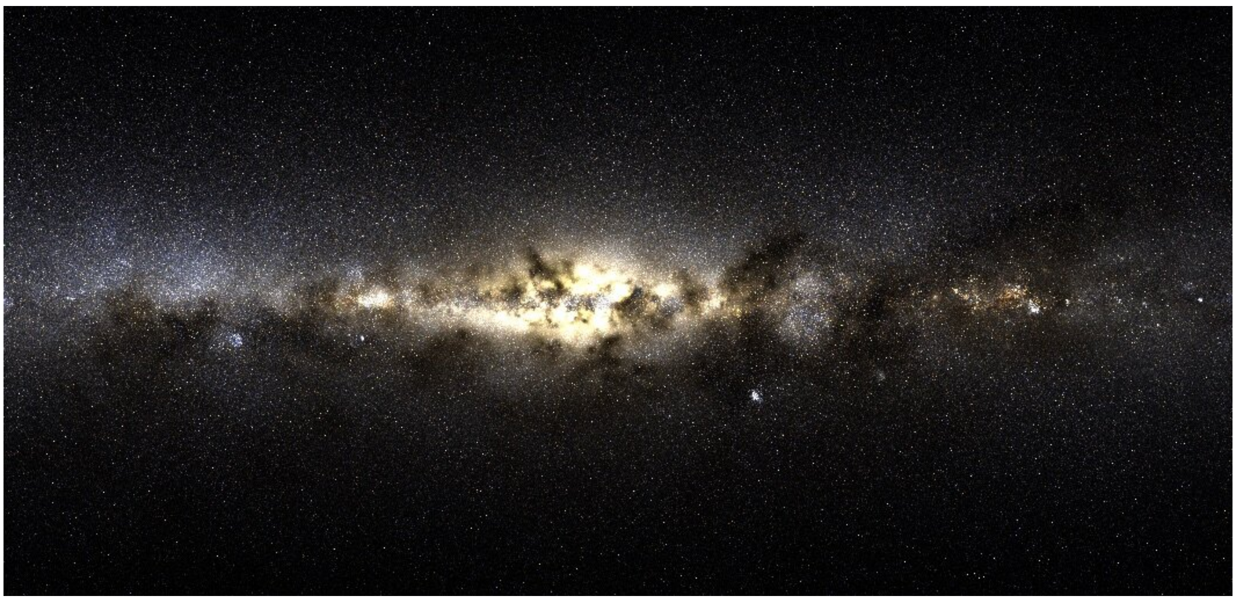
New collection of stars, not born in our galaxy, discovered in Milky Way

July 7 2020, by Aaron Dubrow



Still from a simulation of individual galaxies forming, starting at a time when the Universe was just a few million years old. Credit: Hopkins Research Group, Caltech

Astronomers can go their whole career without finding a new object in the sky. But for Lina Necib, a postdoctoral scholar in theoretical physics at Caltech, the discovery of a cluster of stars in the Milky Way, but not born of the Milky Way, came early—with a little help from supercomputers, the Gaia space observatory, and new deep learning methods.

Writing in *Nature Astronomy* this week, Necib and her collaborators describe Nyx, a vast new stellar stream in the vicinity of the Sun, that may provide the first indication that a dwarf galaxy had merged with the Milky Way disk. These stellar streams are thought to be globular clusters or dwarf galaxies that have been stretched out along its orbit by tidal forces before being completely disrupted.

The discovery of Nyx took a circuitous route, but one that reflects the multifaceted way astronomy and astrophysics are studied today.

FIRE in the Cosmos

Necib studies the kinematics—or motions—of stars and dark matter in the Milky Way. "If there are any clumps of stars that are moving together in a particular fashion, that usually tells us that there is a reason that they're moving together."

Since 2014, researchers from Caltech, Northwestern University, UC San Diego and UC Berkeley, among other institutions, have been developing highly-detailed simulations of realistic galaxies as part of a project called FIRE (Feedback In Realistic Environments). These simulations include everything scientists know about how galaxies form and evolve. Starting from the virtual equivalent of the beginning of time, the simulations produce galaxies that look and act much like our own.

Mapping the Milky Way

Concurrent to the FIRE project, the Gaia space observatory was launched in 2013 by the European Space Agency. Its goal is to create an extraordinarily precise three-dimensional map of about one billion stars throughout the Milky Way galaxy and beyond.

"It's the largest kinematic study to date. The observatory provides the motions of one billion stars," she explained. "A subset of it, seven million stars, have 3-D velocities, which means that we can know exactly where a star is and its motion. We've gone from very small datasets to doing massive analyses that we couldn't do before to understand the structure of the Milky Way."

The discovery of Nyx involved combining these two major astrophysics projects and analyzing them using deep learning methods.

Among the questions that both the simulations and the sky survey address is: How did the Milky Way become what it is today?

"Galaxies form by swallowing other galaxies," Necib said. "We've assumed that the Milky Way had a quiet merger history, and for a while it was concerning how quiet it was because our simulations show a lot of mergers. Now, with access to a lot of smaller structures, we understand it wasn't as quiet as it seemed. It's very powerful to have all these tools, data and simulations. All of them have to be used at once to disentangle this problem. We're at the beginning stages of being able to really understand the formation of the Milky way."

Applying Deep Learning to Gaia

A map of a billion stars is a mixed blessing: so much information, but nearly impossible to parse by human perception.

"Before, astronomers had to do a lot of looking and plotting, and maybe use some clustering algorithms. But that's not really possible anymore," Necib said. "We can't stare at seven million stars and figure out what they're doing. What we did in this series of projects was use the Gaia mock catalogs."

The Gaia mock catalog, developed by Robyn Sanderson (University of Pennsylvania), essentially asked: 'If the FIRE simulations were real and observed with Gaia, what would we see?'

Necib's collaborator, Bryan Ostdiek (formerly at University of Oregon, and now at Harvard University), who had previously been involved in the Large Hadron Collider (LHC) project, had experience dealing with huge datasets using machine and deep learning. Porting those methods over to astrophysics opened the door to a new way to explore the cosmos.

"At the LHC, we have incredible simulations, but we worry that machines trained on them may learn the simulation and not real physics," Ostdiek said. "In a similar way, the FIRE galaxies provide a wonderful environment to train our models, but they are not the Milky Way. We had to learn not only what could help us identify the interesting stars in simulation, but also how to get this to generalize to our real galaxy."

The team developed a method of tracking the movements of each star in the virtual galaxies and labeling the stars as either born in the host galaxy or accreted as the products of galaxy mergers. The two types of stars have different signatures, though the differences are often subtle. These labels were used to train the deep learning model, which was then tested on other FIRE simulations.

After they built the catalog, they applied it to the Gaia data. "We asked the neural network, 'Based on what you've learned, can you label if the stars were accreted or not?'" Necib said.

The model ranked how confident it was that a star was born outside the Milky Way on a range from 0 to 1. The team created a cutoff with a tolerance for error and began exploring the results.

This approach of applying a model trained on one dataset and applying it to a different but related one is called transfer learning and can be fraught with challenges. "We needed to make sure that we're not learning artificial things about the simulation, but really what's going on in the data," Necib said. "For that, we had to give it a little bit of help and tell it to reweigh certain known elements to give it a bit of an anchor."

They first checked to see if it could identify known features of the galaxy. These include "the Gaia sausage"—the remains of a dwarf galaxy that merged with the Milky Way about six to ten billion years ago and that has a distinctive sausage-like orbital shape.

"It has a very specific signature," she explained. "If the neural network worked the way it's supposed to, we should see this huge structure that we already know is there."

The Gaia sausage was there, as was the stellar halo—background stars that give the Milky Way its tell-tale shape—and the Helmi stream, another known dwarf galaxy that merged with the Milky Way in the distant past and was discovered in 1999.

First Sighting: Nyx

The model identified another structure in the analysis: a cluster of 250 stars, rotating with the Milky Way's disk, but also going toward the center of the galaxy.

"Your first instinct is that you have a bug," Necib recounted. "And you're like, 'Oh no!' So, I didn't tell any of my collaborators for three weeks. Then I started realizing it's not a bug, it's actually real and it's new."

But what if it had already been discovered? "You start going through the literature, making sure that nobody has seen it and luckily for me, nobody had. So I got to name it, which is the most exciting thing in astrophysics. I called it Nyx, the Greek goddess of the night. This particular structure is very interesting because it would have been very difficult to see without machine learning."

The project required advanced computing at many different stages. The FIRE and updated FIRE-2 simulations are among the largest computer models of galaxies ever attempted. Each of the nine main simulations—three separate galaxy formations, each with slightly different starting point for the sun—took months to compute on the largest, fastest supercomputers in the world. These included Blue Waters at the National Center for Supercomputing Applications (NCSA), NASA's High-End Computing facilities, and most recently Stampede2 at the Texas Advanced Computing Center (TACC).

The researchers used clusters at the University of Oregon to train the deep learning model and to apply it to the massive Gaia dataset. They are currently using Frontera, the fastest system at any university in the world, to continue the work.

"Everything about this project is computationally very intensive and would not be able to happen without large-scale computing," Necib said.

Future Steps

Necib and her team plan to explore Nyx further using ground-based telescopes. This will provide information about the chemical makeup of the stream, and other details that will help them date Nyx's arrival into the Milky Way, and possibly provide clues on where it came from.

The next data release of Gaia in 2021 will contain additional information about 100 million stars in the catalog, making more discoveries of accreted clusters likely.

"When the Gaia mission started, astronomers knew it was one of the largest datasets that they were going to get, with lots to be excited about," Necib said. "But we needed to evolve our techniques to adapt to the dataset. If we didn't change or update our methods, we'd be missing out on physics that are in our dataset."

The successes of the Caltech team's approach may have an even bigger impact. "We're developing computational tools that will be available for many areas of research and for non-research related things, too," she said. "This is how we push the technological frontier in general."

More information: Lina Necib et al, Evidence for a vast prograde stellar stream in the solar vicinity, *Nature Astronomy* (2020). DOI: 10.1038/s41550-020-1131-2

Provided by University of Texas at Austin