[Left side]
Illustration of the exoplanet LHS 3844 b. It is a rocky planet about 1.3 times Earth's size located about 49 light-years away in the constellation Indus, making it among the closest transiting exoplanets known. The star is a cool M-type dwarf star about one-fifth the size of our Sun. Completing an orbit every 11 hours, the planet lies so close to its star that some of its rocky surface on the daytime side may form pools of molten lava.

[Right side]
The NAS supercomputer cluster supports many key NASA programs and research areas, from modeling airflow over the blades of helicopters, to simulating future global climate trends for our changing planet, to providing critical analysis for design decisions about the Space Launch System rocket that will someday take humans to Mars.

This mosaic of the northern sky was assembled from 208 images taken by NASA's Transiting Exoplanet Survey Satellite (TESS) during its second year of science operations, completed in July 2020. The mission split the northern sky into 13 sectors, each of which was imaged for nearly a month by the spacecraft's four cameras. Among the many notable celestial objects visible: the glowing arc and obscuring dust clouds of the Milky Way (left), our home galaxy seen edgewise; the Andromeda galaxy (oval, center left), our nearest large galactic neighbor, located 2.5 million light-years away; and the North America Nebula (lower left), part of a stellar factory complex 1,700 light-years away. The prominent dark lines are gaps between the detectors in TESS's camera system.

[Top right]
Artist concept of the TESS spacecraft.

The locations of nearly 1,000 candidate exoplanets identified by Sept. 1, 2019, are plotted on the TESS mosaic (orange dots). Astronomers are studying these targets to confirm new worlds. The host stars of the 29 TESS planet discoveries to date (2019) are shown on the mission's southern sky mosaic (cyan circles).

The Vetting Challenge: Finding the Signal in the Noise

While the goal is to find the clean, U-shaped transit of a true exoplanet (center), the raw data is flooded with astrophysical and instrumental imposters. To automate discovery at scale, our models must be trained to reject eclipsing binaries (top-left), stellar activity (bottom-left), foreground objects like comets (top-right), and chaotic spacecraft hardware noise (bottom-right).

The Data Pipeline: From Big Picture to Transit Signal

To find a planet, we must isolate a microscopic signal from a massive dataset. We begin with the crowded Full Frame Image (left), identify a specific target's postage stamp of pixels (right), and extract a 1D light curve. By processing this time series data and evaluating spatial centroid shifts (bottom center), processing pipelines like TESS SPOC at NASA Ames phase-fold the data (bottom left) to reveal the hidden planetary transit.

ExoMiner: The Deep Learning Classifier

ExoMiner acts as a highly consistent, automated expert. It utilizes a multi-branch Convolutional Neural Network (CNN) to simultaneously process 1D temporal data (phase-folded light curves), 2D spatial features (difference images), periodogram, and scalar diagnostic tests from the TESS SPOC Data Validation pipeline. By concatenating these inputs, the model accurately distinguishes true planets from astrophysical false positives, outputting a high-precision confidence score in a fraction of a second.

Scaling Inference on NASA HECC

ExoMiner is deployed as a highly portable, cross-platform container via GitHub. Utilizing rootless Podman on NASA's Cabeus supercomputer, the pipeline leverages Hopper H100 GPUs to train on tens of thousands of signals. For large-scale inference, the CPU-optimized architecture allows the model to instantly scale across idle 72-core GH200 nodes, rapidly evaluating hundreds of thousands of light curves in parallel.

The Science Yield: Validating Hidden Worlds

By running ExoMiner at scale, we transformed raw telescope data into a massive scientific yield. The pipeline has officially validated hundreds of new exoplanets from the Kepler (top left) and is now aiding exoplanet scientists vetting transit-like signals from TESS. As humanity celebrates the 1,000th TESS exoplanet discovery milestone this month, the speed and accuracy of ML vetting are proving critical to expanding our statistical understanding of the galaxy.

The Future: Completing the Exoplanet Census

While Kepler and TESS mastered the transit method for close-in planets, NASA's next flagship missions will push the boundaries of data science. The Roman Space Telescope will search the dense Galactic Bulge using microlensing, generating over a terabyte of data daily. Meanwhile, the Habitable Worlds Observatory will rely on extreme image processing to directly image Earth-analogs. Meeting these new computational challenges will require the continued evolution of High-End Computing, scalable pipelines, and highly capable ML models.

Full Presentation Acknowledgments & Visual Credits

NASA Centers & Teams: NASA Goddard Space Flight Center (SVS, CI Lab), NASA Ames Research Center (NAS), the NASA/MIT/TESS Team, and the NASA Exoplanet Archive.

Institutions & Tech: KBR Inc., The SETI Institute, USRA, UMCP, ADNET Systems, SGT, Maslow Media Group, GitHub, and Podman.

Individual Creators, Scientists & Producers: Scott Wiessinger, Francis Reddy, Claire Andreoli, Max Petruning, Ethan Kruse, Aaron E. Lepsch, Britt Griswold, Chris Smith, Joy Ng, Kathalina Tran, Karen Fox, Chris Henze, Danielle Futselaar, and Veselin Kostov.