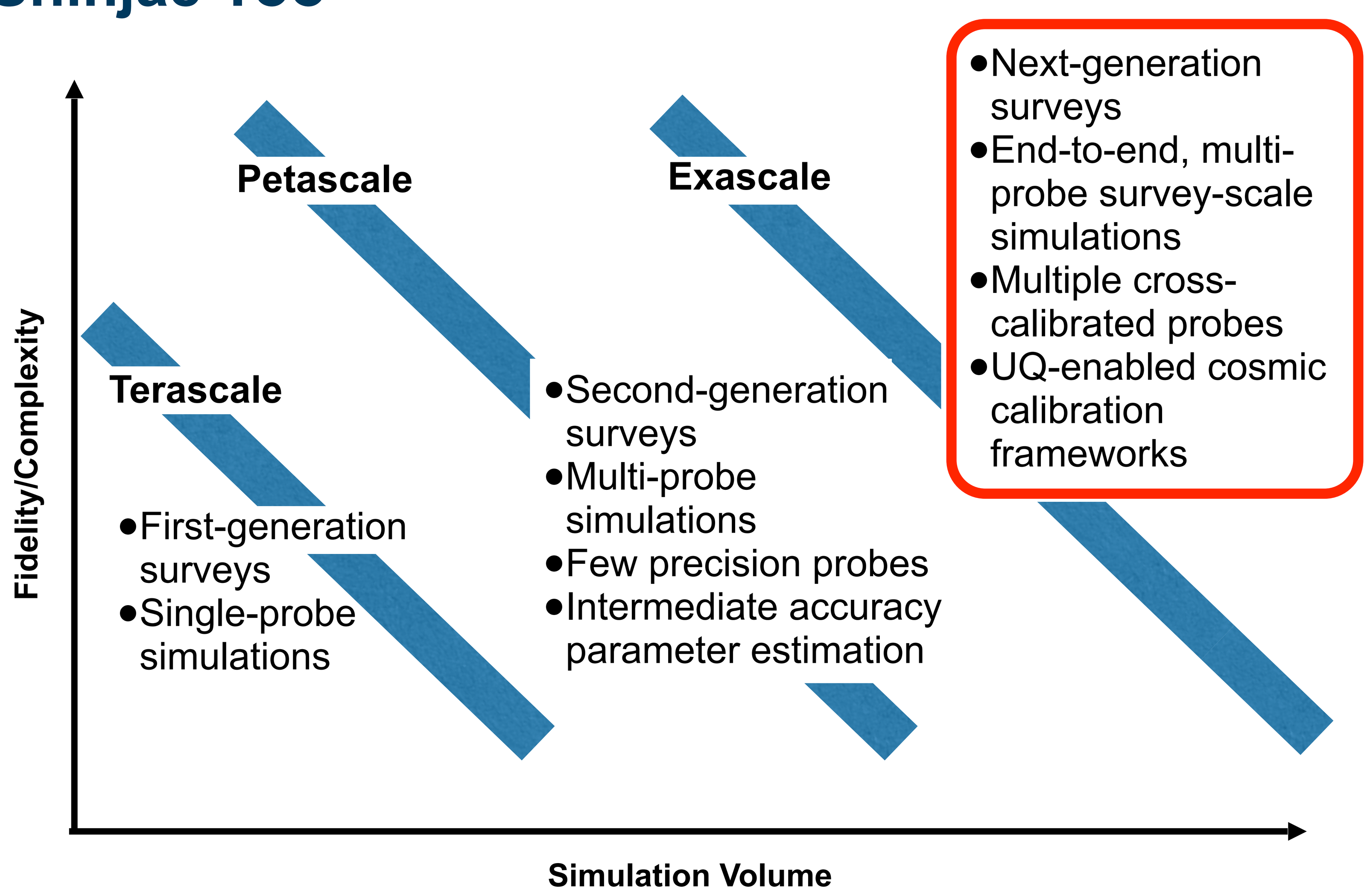
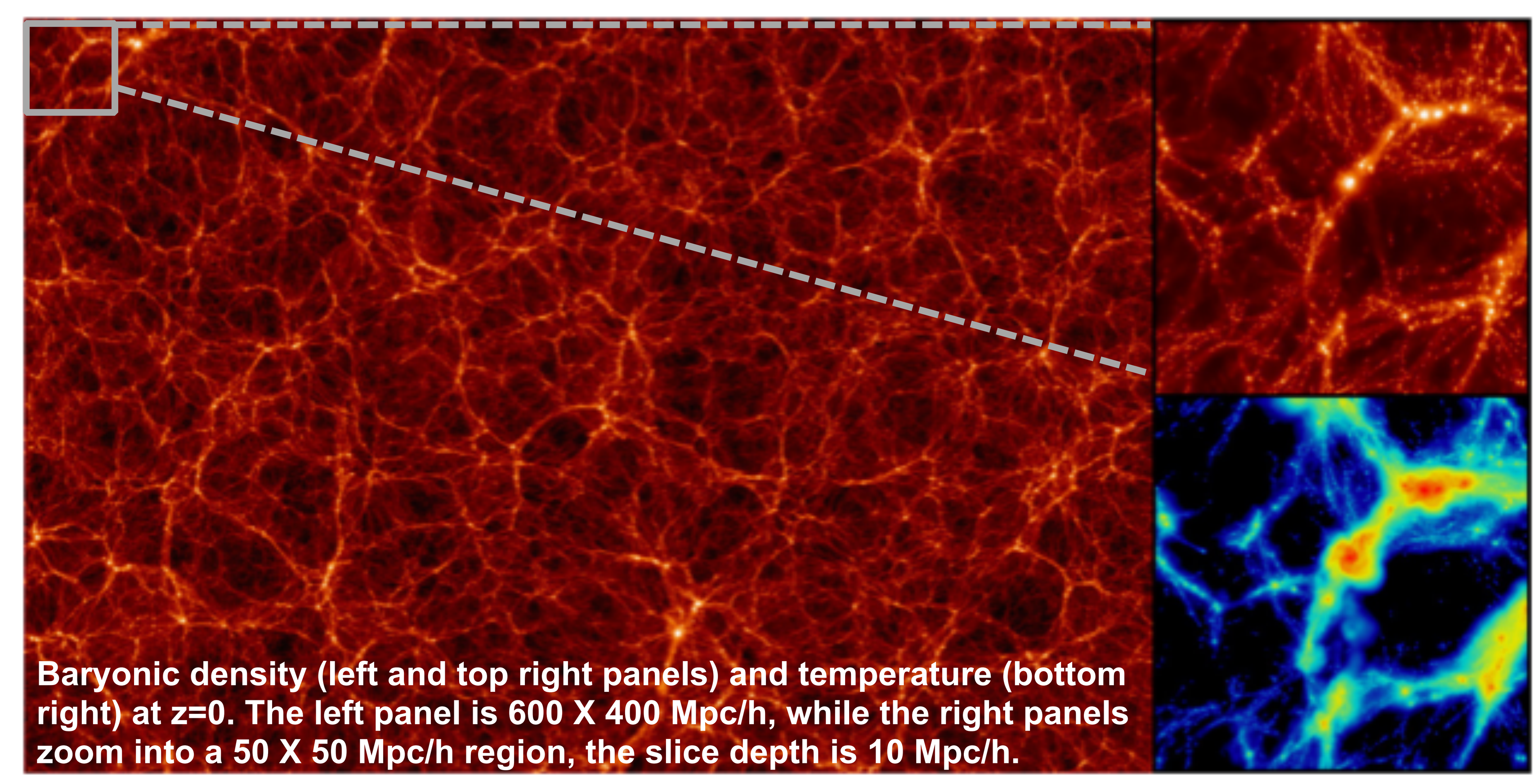
Accelerating HEP Science: Inference and Machine Learning at Extreme Scales

Prasanna Balaprakash, Mickael Binois, Arindam Fadikar, Robert Gramacy, Salman Habib, Katrin Heitmann, David Higdon, Earl Lawrence, Yuewei Lin, Zarija Lukic, Dimitriy Morozov, Nesar Ramachandra, Anze Slosar, Stefan Wild, Shinjae Yoo

BROOKHAVEN NATIONAL LABORATORY
Argonne NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY EST. 1943
VIRGINIA TECH.

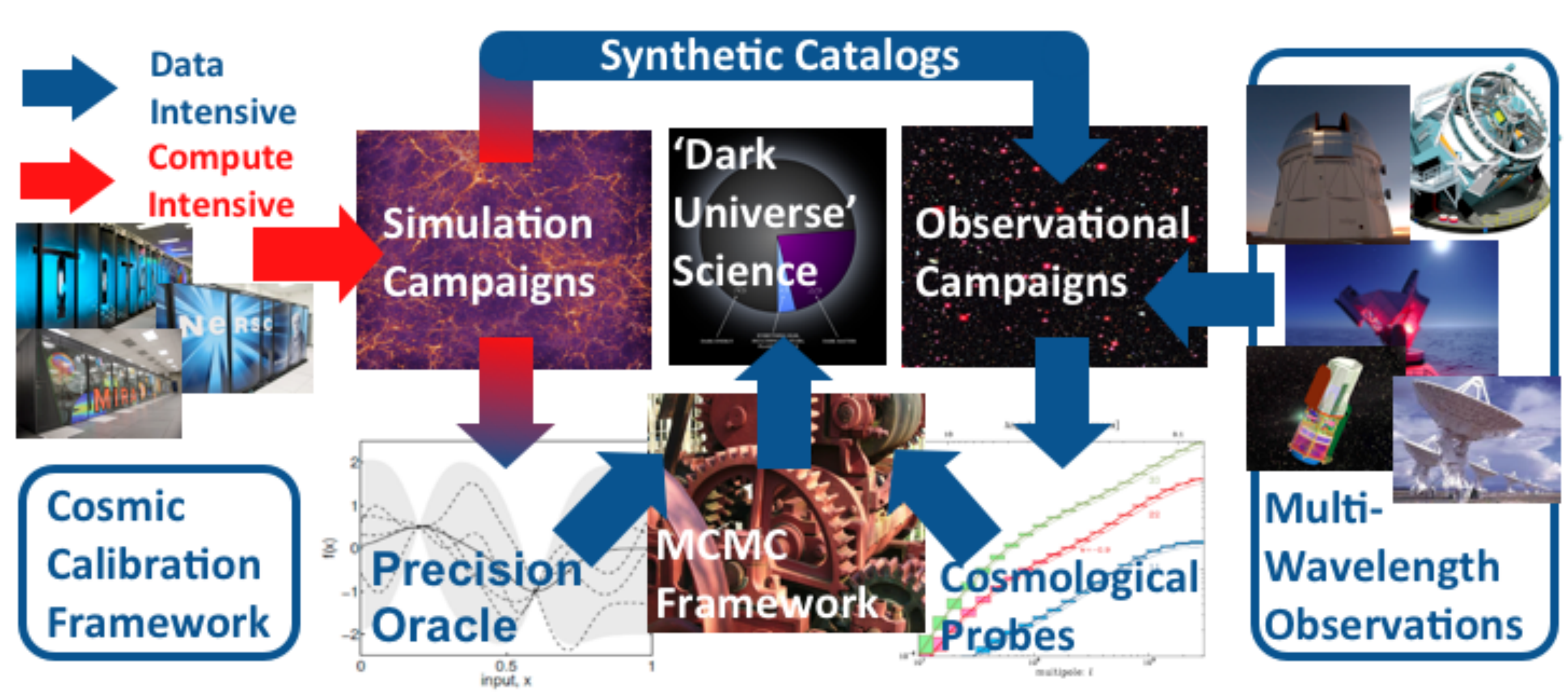


- **Simulations and Sky Surveys**
  - **Sky surveys** can measure certain cosmological statistics (including cross-correlations) at the 1% level or better; scientific inference from these measurements is strongly limited by shortcomings in theory and modeling
  - **Exascale simulations**, constructed from improved theory and modeling approaches (including subgrid modeling), implemented with new algorithms, and melded with advanced V&V and UQ protocols have the potential to fill current simulation gaps for a number of cosmological probes



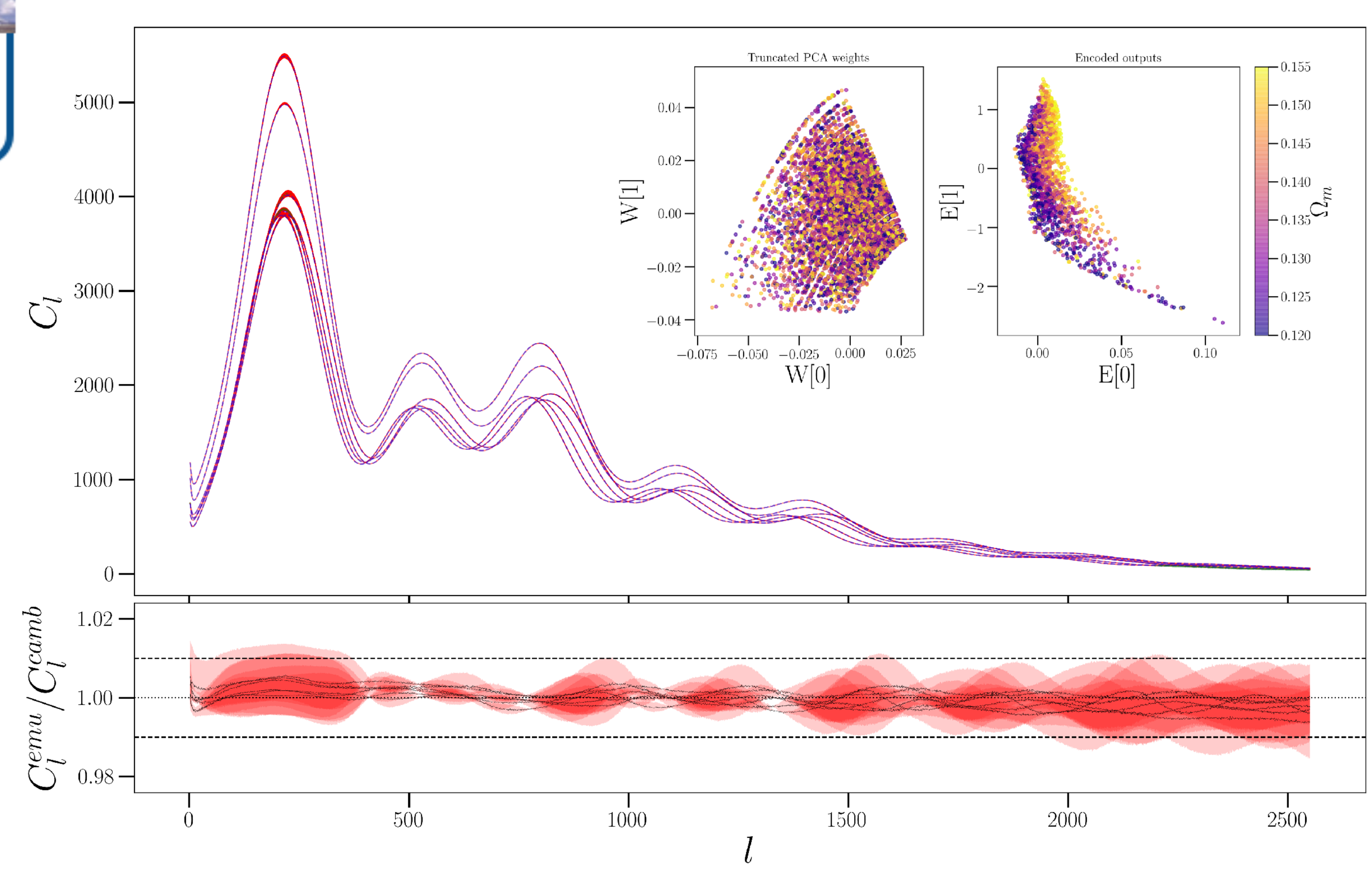
Baryonic density (left and top right panels) and temperature (bottom right) at z=0. The left panel is 600 X 400 Mpc/h, while the right panels zoom into a 50 X 50 Mpc/h region, the slice depth is 10 Mpc/h.

BorgCube: First large-scale CRK-HACC simulation containing 2 X 2304^3 dark matter plus baryonic particles in a cubic volume of side length 800 Mpc/h.

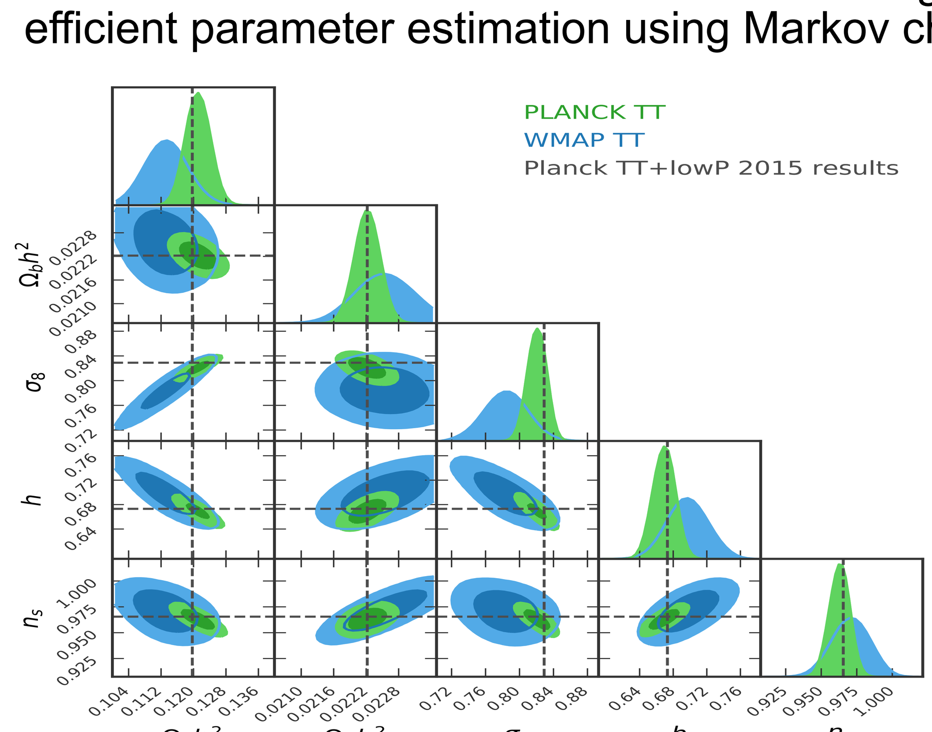


Cosmological Scientific Inverse Problem: Data and information flow in a precision cosmology application using emulators built using modern statistical and machine learning methods; these techniques can also be used to efficiently generate synthetic galaxy catalogs.

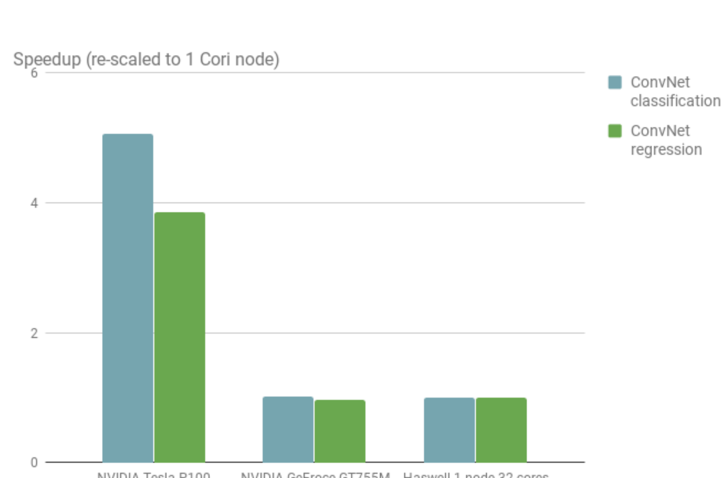
- **Emulation using Variational Autoencoders**
  - **Dimensional Reduction** is used in constructing emulators to make Gaussian process-based interpolation numerically tractable; traditionally this has been done using Principal Component Analysis (PCA). In initial work, we have shown that accurate emulators can be constructed using autoencoders; the resulting emulators are ~2000 times faster than the original code, allowing for very efficient parameter estimation using Markov chain Monte Carlo



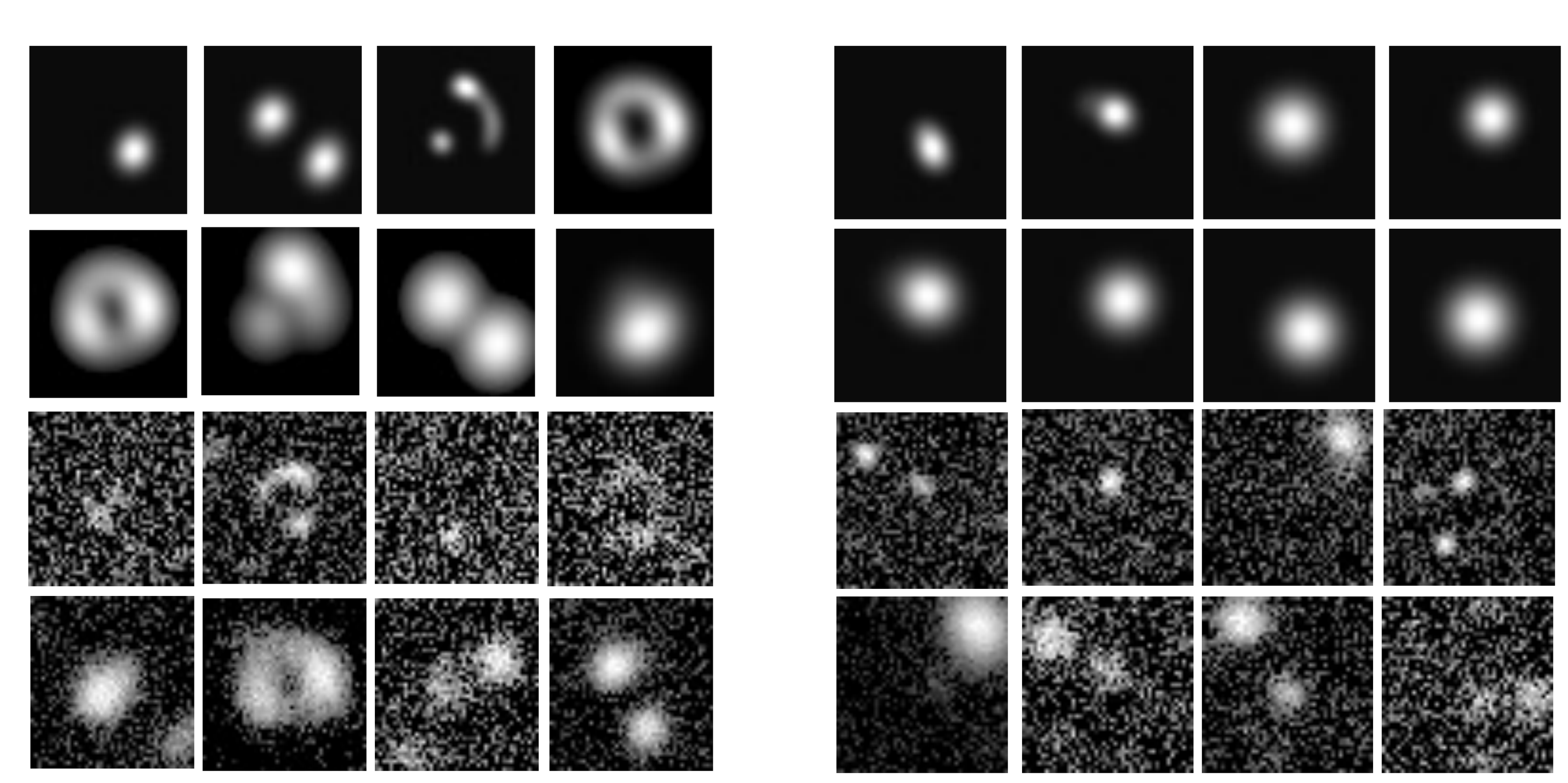
Cosmic Microwave Background Emulation: Angular power spectrum for multiple cosmologies with 5 free parameters, exact and emulated spectra superimposed; dimensional reduction was perfromed using a variational autoencoder. The 1% error behavior is evident in the lower panel. The insets show a comparison of a 2-D PCA dimensional reduction (left) versus an autoencoder (right).



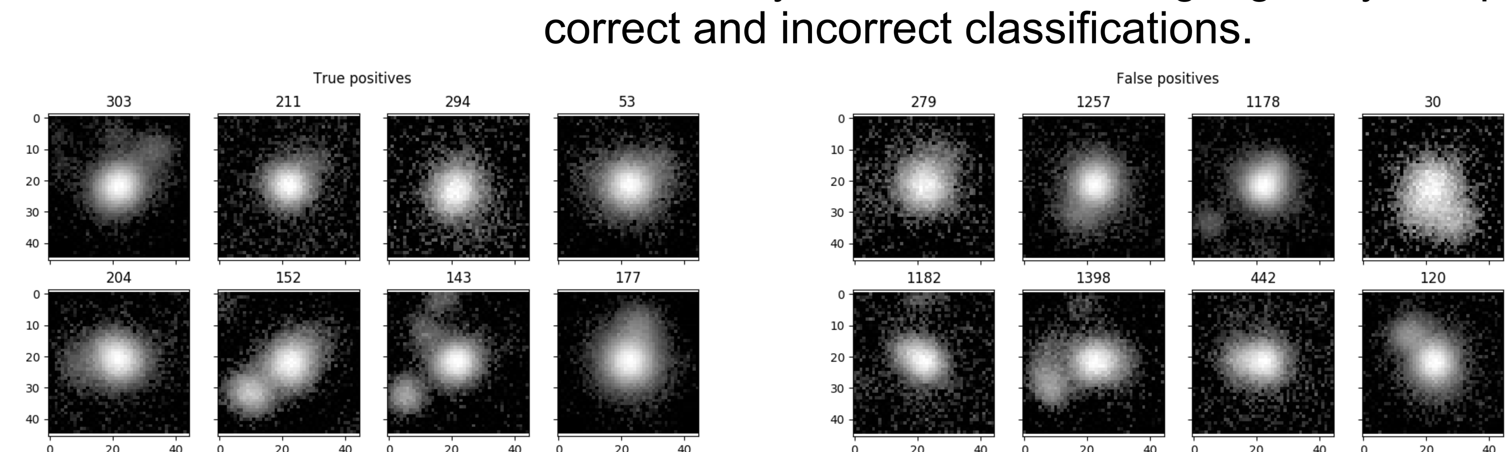
Cosmological Parameters: Emulator-based MCMC results for cosmological parameters from WMAP and Planck satellite observations of the CMB thermal anisotropy power spectrum. Emulators have been constructed for the polarization and cross-power spectra and wil be used for analysis of South Pole Telescope (SPT) datasets.



Strong Lensing Image Classification: Strongly-lensed background galaxies have distorted shapes (arc-like to rings) but these are hard to discern in the presence of noise and can be confused with non-lensed background objects — with billions of candidates in future surveys, the task is beyond human capability. A CNN-based classifier was developed using noisy simulated images, achieving 80-90% accuracy with a classifcation time of only 10 microseconds per image using state-of-the-art GPUs. The difficulty of the task can be gauged by comparing correct and incorrect classifications.



Strong Lensing Image Classification: Left — lensed examples, top two rows (noiseless), bottom two rows with sky noise; right — coreesponding unlensed images. Convolutional neural networks were trained on a large number of simulated images for fast classification tasks.



Correct and Incorrect Image Classification: Left — correct classifications, right incorrect.