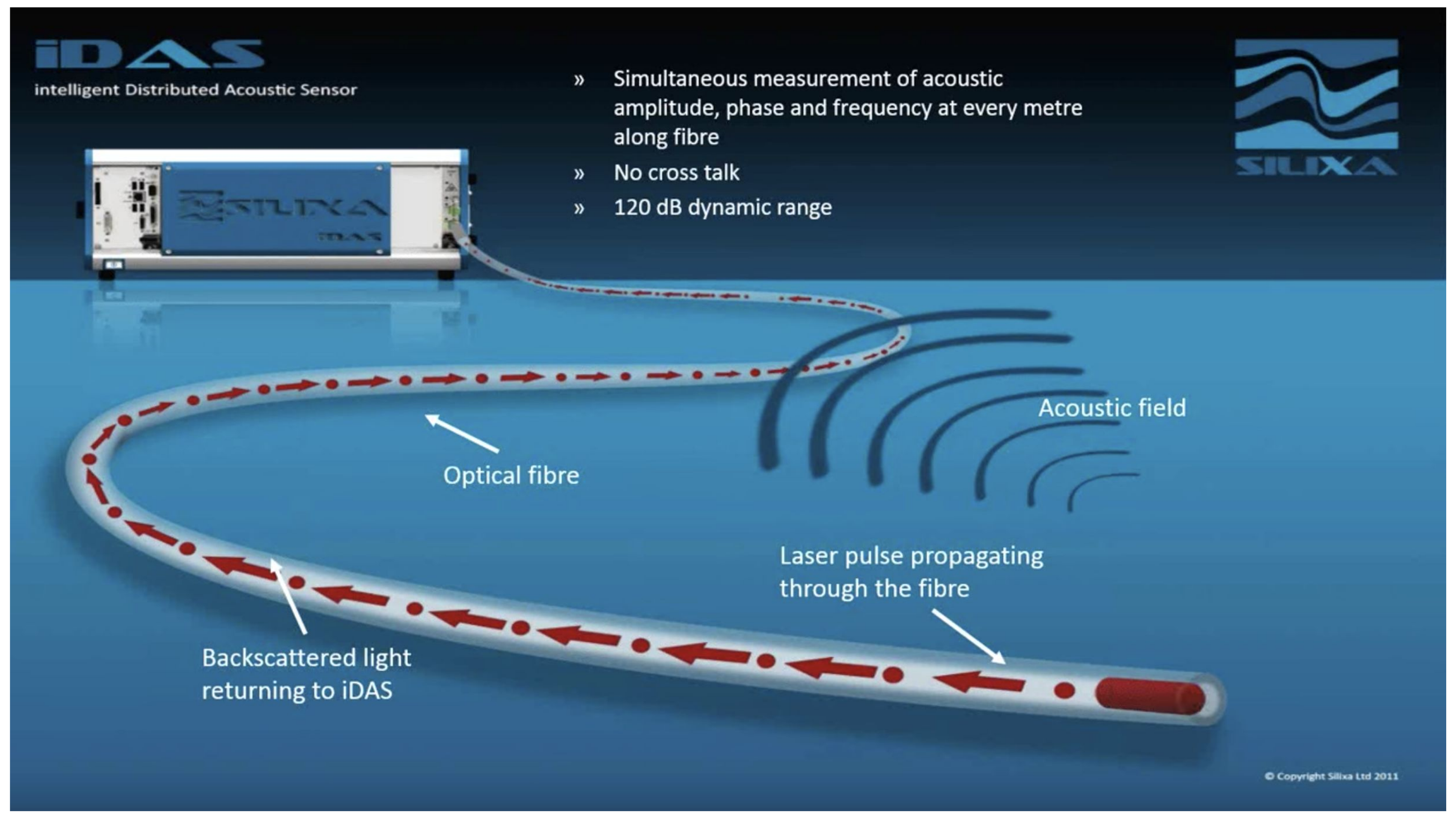
Deep Learning on Real Geophysical Data: A Case Study for Distributed Acoustic Sensing Research

V. Dumont[1], V. Rodríguez Tribaldos[1], J. Kesheng Wu[1] and J. Ajo-Franklin[2]

[1] Lawrence Berkeley National Laboratory, [2] Rice University

Introduction



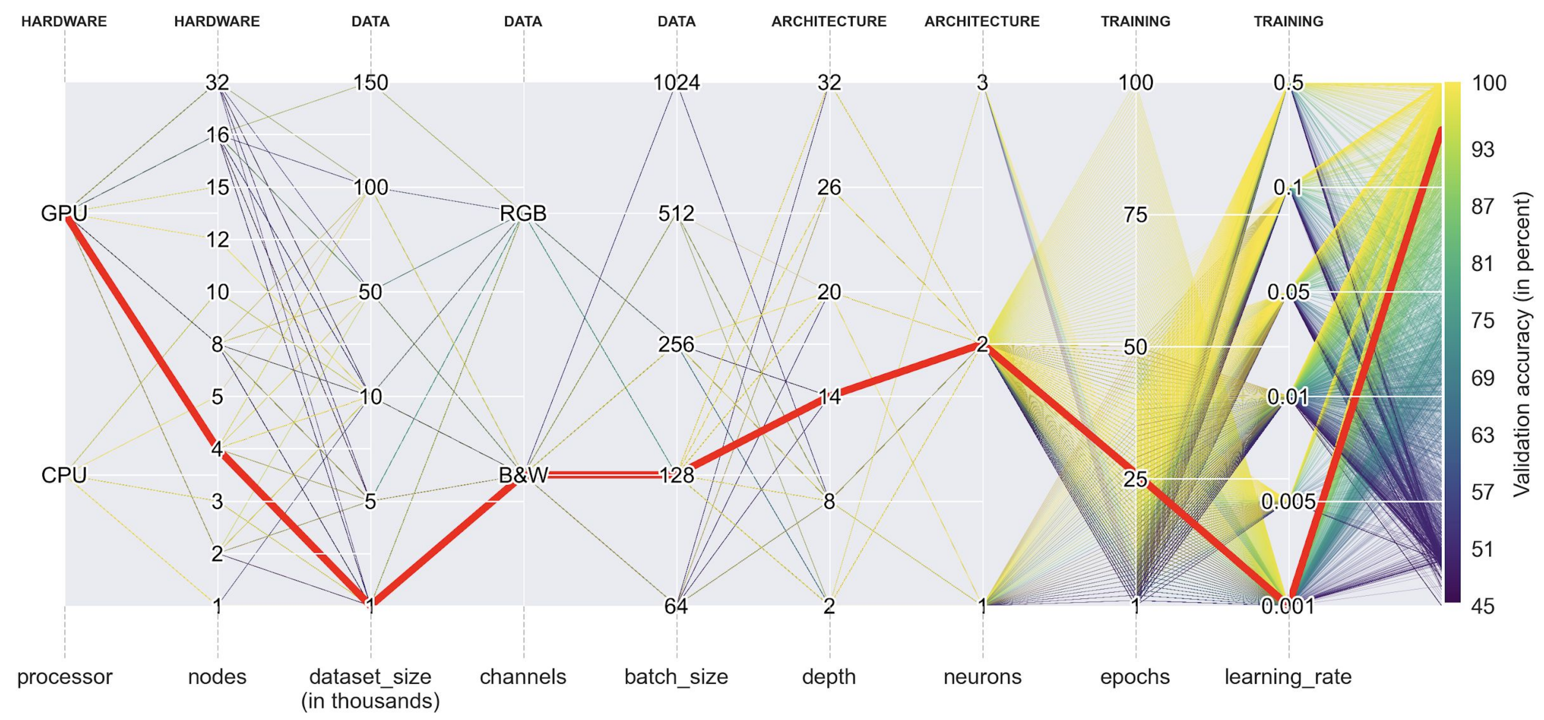
Distributed Acoustic Sensing (DAS) is an emerging technology that repurposes commercial fiber-optic cables as strain-rate sensing arrays. DAS can be deployed on existing telecommunication networks not currently used for data transfer, known as dark fiber networks (i.e. unlit). Thanks to this technology, DAS enables the acquisition of high-resolution (1m) seismic datasets across tens of kilometers in a variety of environments. This high-density of measurements results in very large datasets that can amount to several terabytes per day, making it very challenging for geophysicists to handle and process them.

DAS arrays can record so-called surface seismic waves generated by many sources, including moving vehicles, trains, etc. These signals can be processed and analyzed to obtain information on near-surface structure by applying a seismological approach called ambient noise interferometry (Ajo-Franklin et al, 2019). Currently, extracting this useful seismic energy from these massive, noisy datasets is very challenging.

Can deep learning be used to identify and extract usable surface waves for optimizing the ambient noise interferometry workflow?

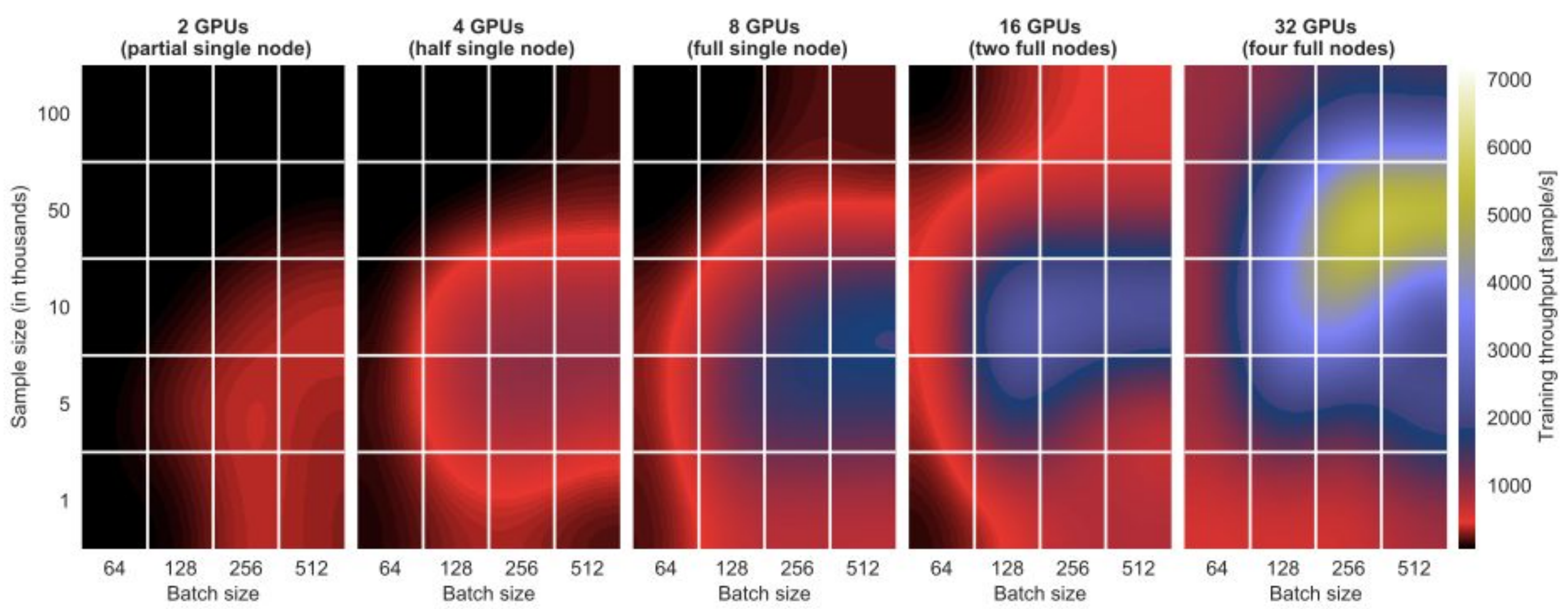
Hyperparameter Tuning

A 9-dimensional parameter space was explored where each parameter evaluates different aspects of the training, including HPC hardware, data characteristics and neural network architecture. A total of 6,460 different trained models are presented.



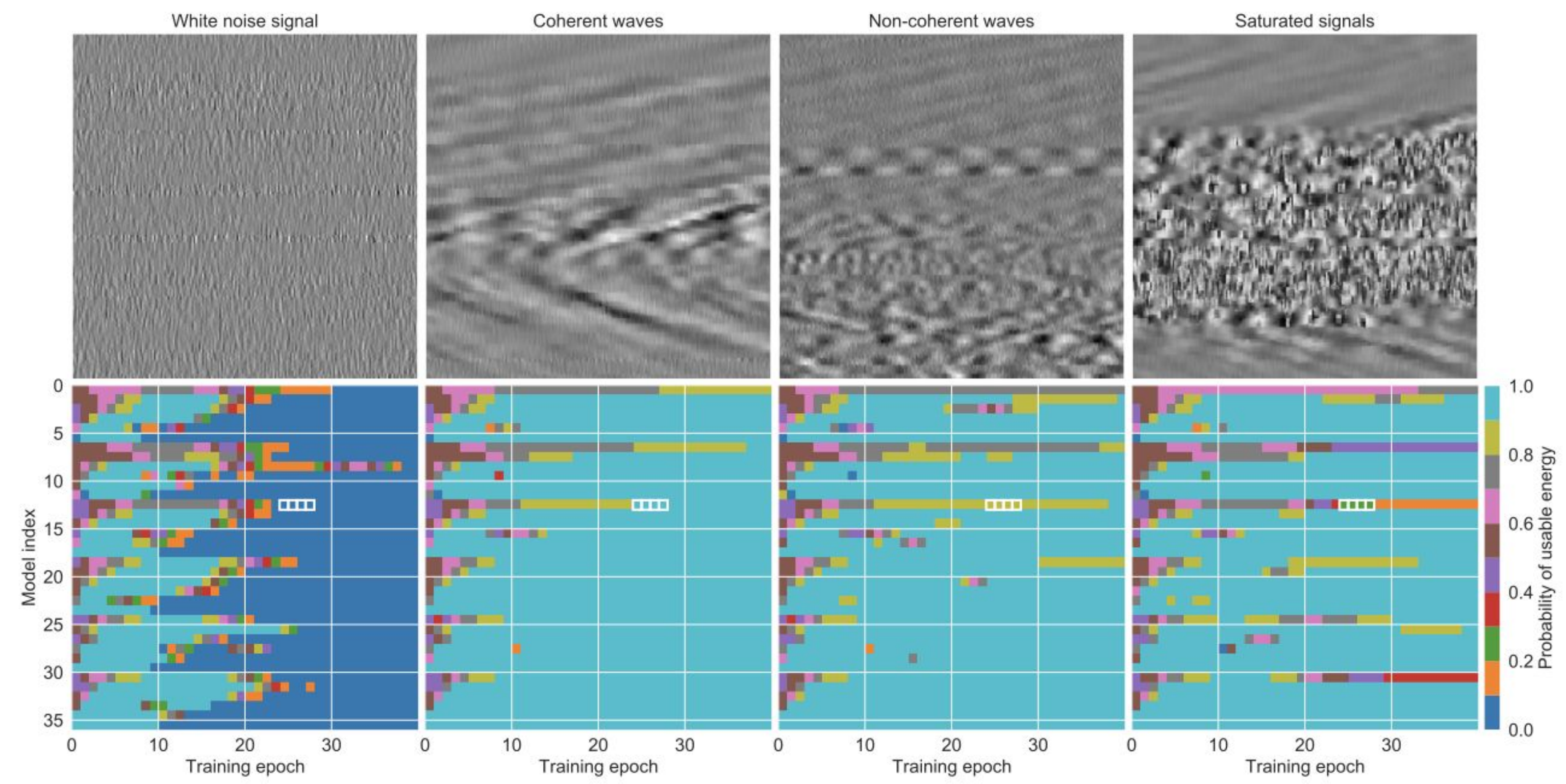
Superlinear Speedup

The combination of data set size and batch size providing the fastest training will change with the number of GPUs used. We were able to improve the throughput by more than two orders of magnitude on a 50,000-image data set trained with a batch size of 512. Indeed, a training speed of 65 sample/second was achieved using 2 GPUs while we managed to reach a throughput of 7,210 sample/second when using a 32 GPUs distributed setup.



Results

We visualized the output probabilities for usable surface wave energies for four reference DAS data regions over the entire set of trained models. The output values were compared with the estimated ground truth for each of those region so as to determine which trained model provide a sensible, trustworthy estimation for usable energy signals.



While the models were not trained on contaminated (i.e., non-coherent, saturated) data regions, we were able to select the best classifier among thousands of trained models using the expected probability range on only four reference images.

Acknowledgment

This work is supported by the Laboratory Directed Research and Development (LDRD) Program of LBNL under U.S. Department of Energy Contract No. DE-AC02-05CH11231. It used resources of the National Energy Research Scientific Computing Center and Energy Science network (ESnet), both are funded under the above contract.