SCEC Southern California Earthquake Center
AN NSF+USGS CENTER

Seismological Laboratory
California Institute of Technology

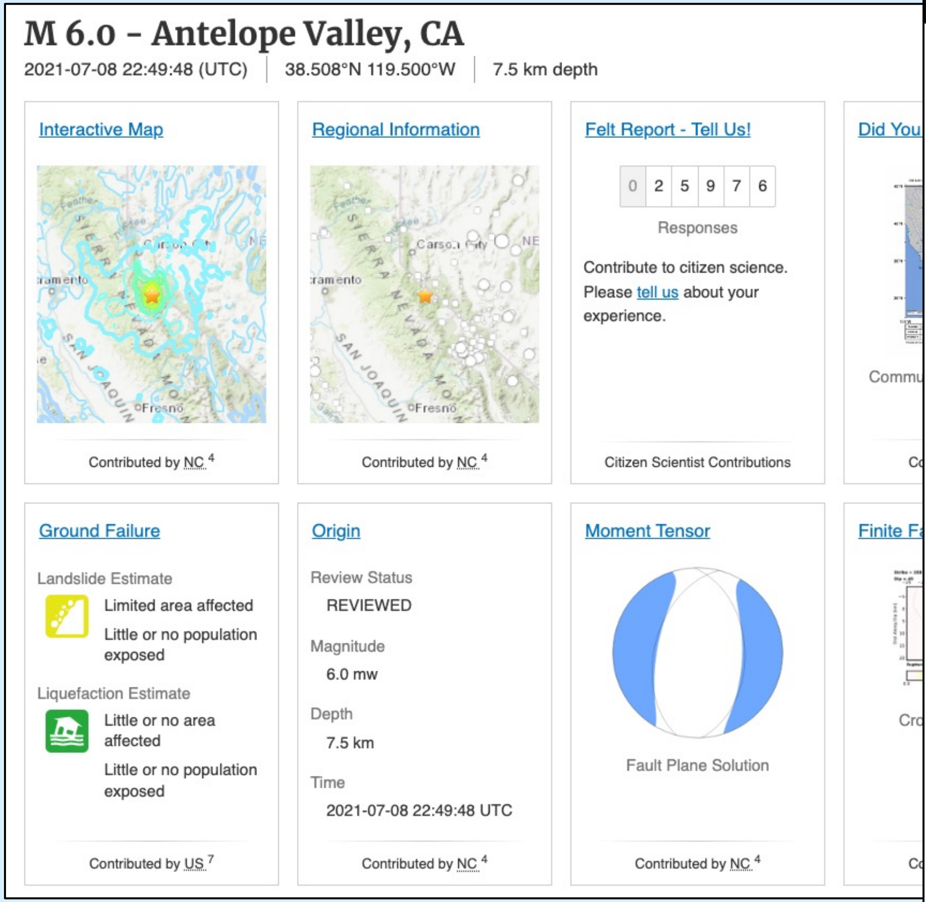
SCEC Annual Meeting:
Session 7: Computational Earthquake Science

Challenges, opportunities, and discoveries using large-scale distributed acoustic sensing arrays

Ettore Biondi*, Yan Yang, Jiaqi Fang, Jiuxun Yin,
Weiqiang Zhu, Jiaxuan Li, Ethan F. Williams,
Zhongwen Zhan

April 14th, 2022

Conventional stations and DAS



M 6.0 - Antelope Valley, CA

2021-07-08 22:49:48 (UTC) | 38.508°N 119.500°W | 7.5 km depth

Interactive Map

Contributed by NC [4]

Regional Information

Contributed by NC [4]

Felt Report - Tell Us!

0 2 5 9 7 6

Responses

Contribute to citizen science. Please tell us about your experience.

Citizen Scientist Contributions

Did You

Commu

Ground Failure

Landslide Estimate

Limited area affected

Little or no population exposed

Liquefaction Estimate

Little or no area affected

Little or no population exposed

Contributed by US [7]

Origin

Review Status: REVIEWED

Magnitude: 6.0 mw

Depth: 7.5 km

Time: 2021-07-08 22:49:48 UTC

Contributed by NC [4]

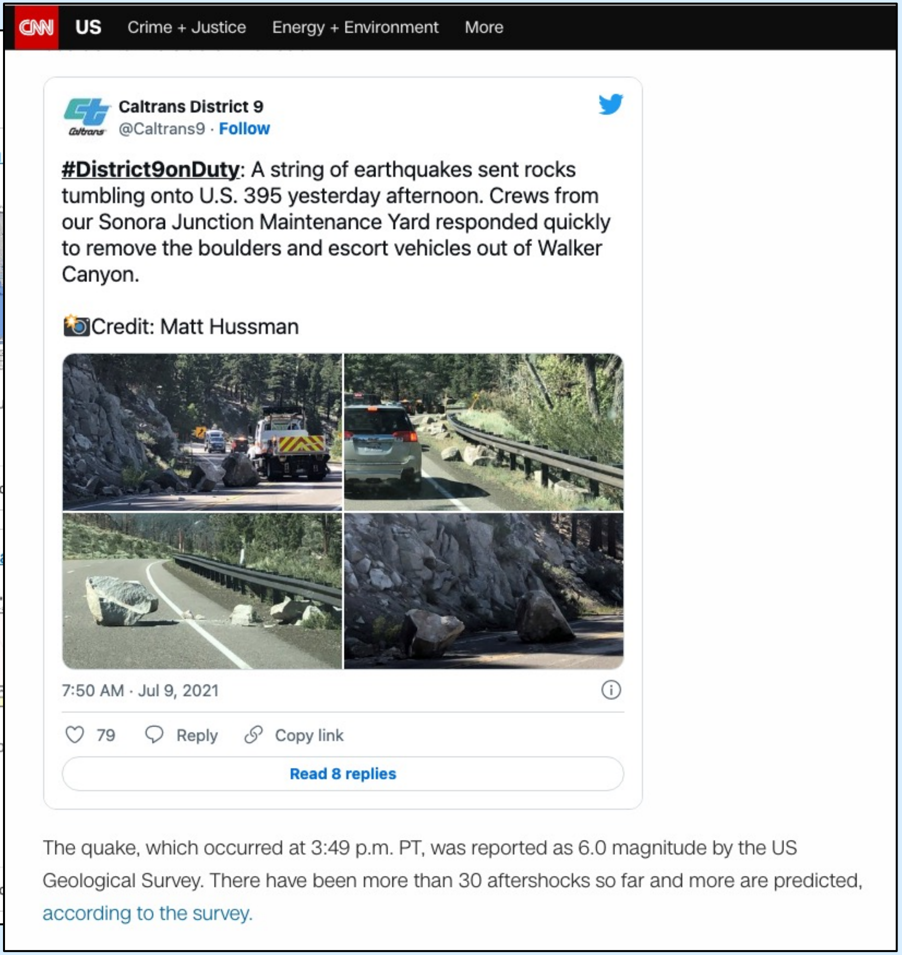
Moment Tensor

Fault Plane Solution

Contributed by NC [4]

Finite Fa

Cro



Caltrans District 9
@Caltrans9 · Follow

#District9onDuty: A string of earthquakes sent rocks tumbling onto U.S. 395 yesterday afternoon. Crews from our Sonora Junction Maintenance Yard responded quickly to remove the boulders and escort vehicles out of Walker Canyon.

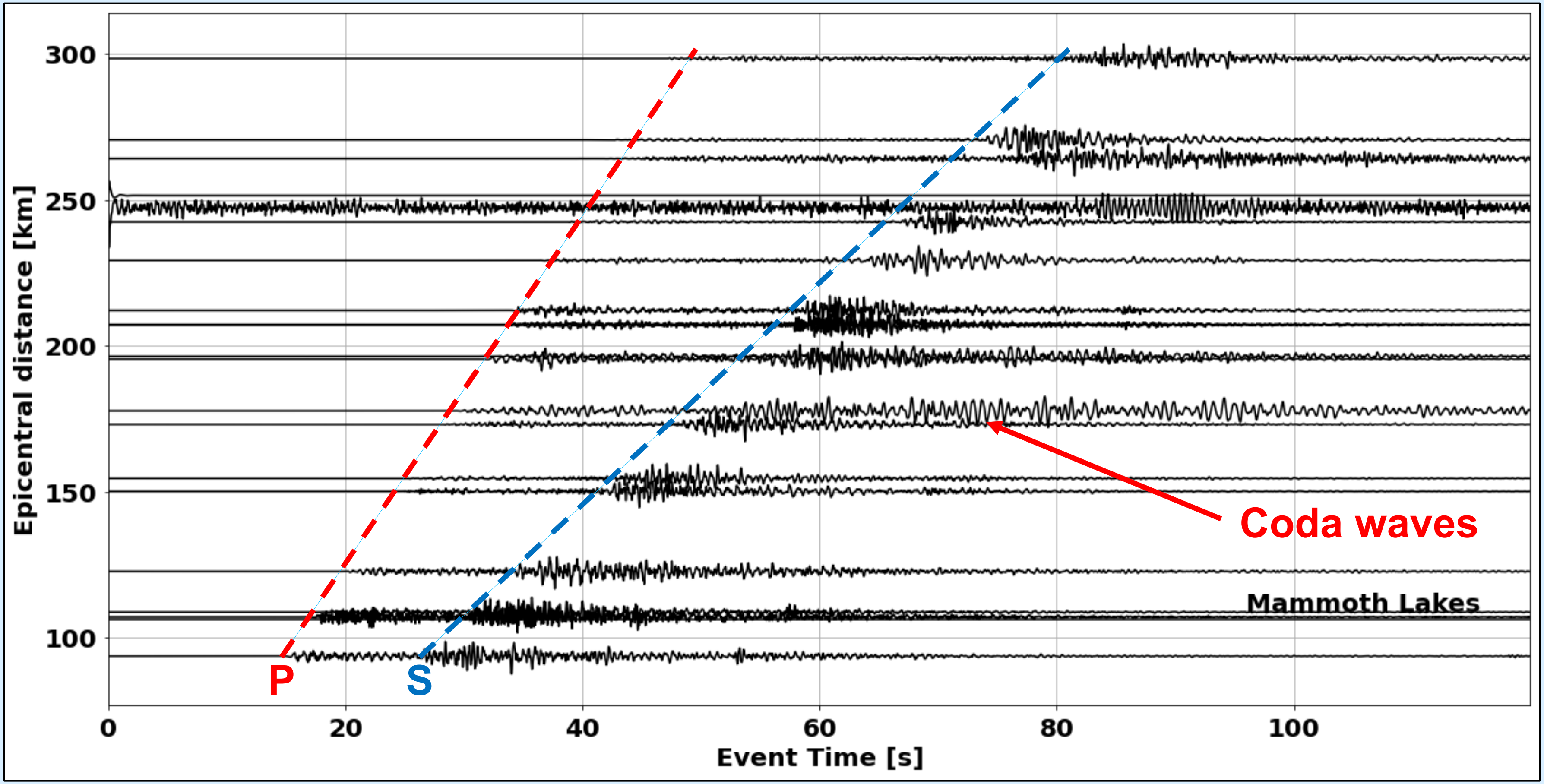
📷Credit: Matt Hussman

7:50 AM · Jul 9, 2021

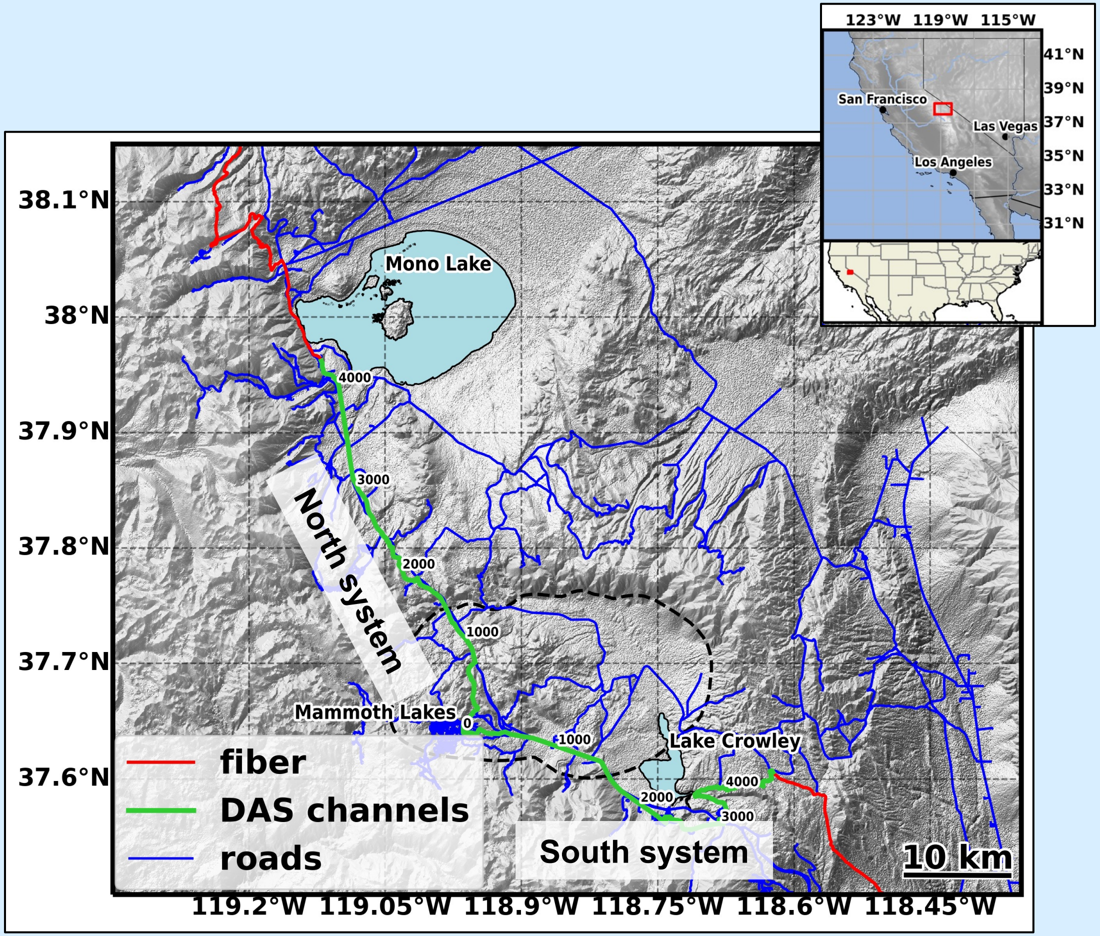
The quake, which occurred at 3:49 p.m. PT, was reported as 6.0 magnitude by the US Geological Survey. There have been more than 30 aftershocks so far and more are predicted, according to the survey.

USGS

Conventional stations and DAS

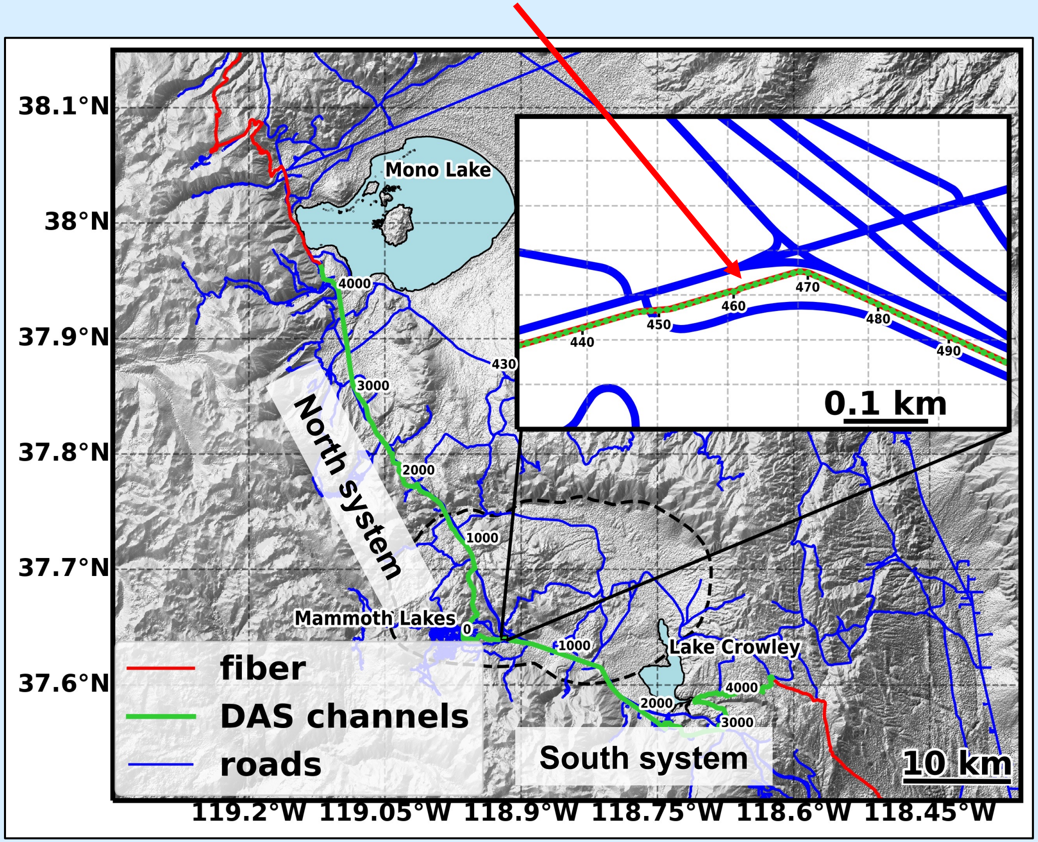


Conventional stations and DAS

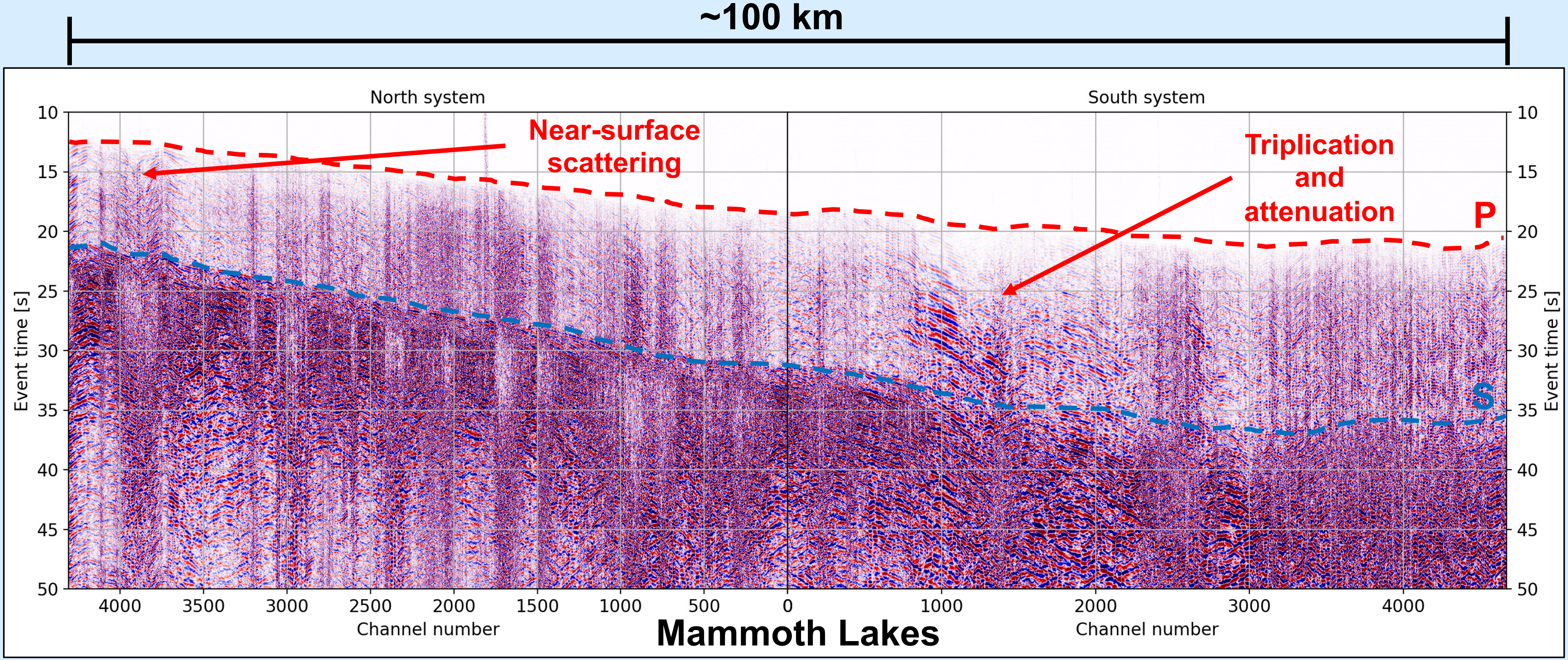


Conventional stations and DAS

DAS provides an ultra-dense array!



Conventional stations and DAS



Conventional stations and DAS

What is the catch?

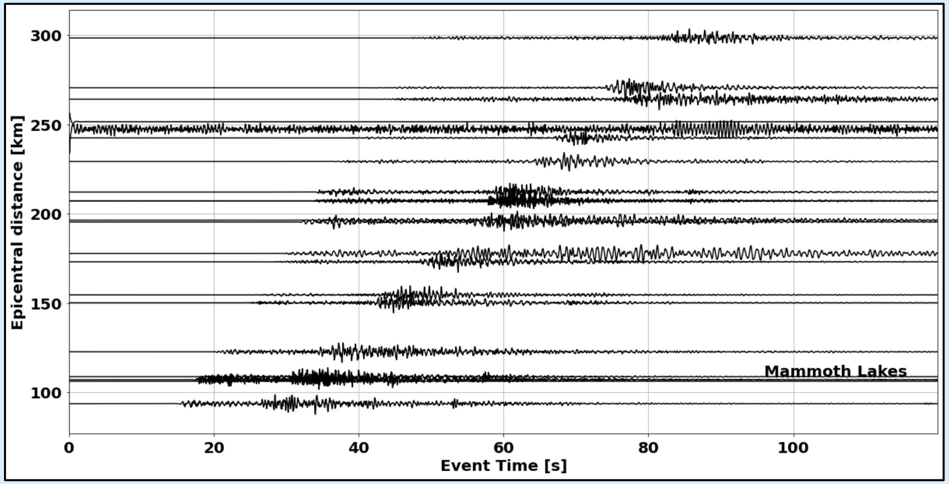
60s @ 100 Hz

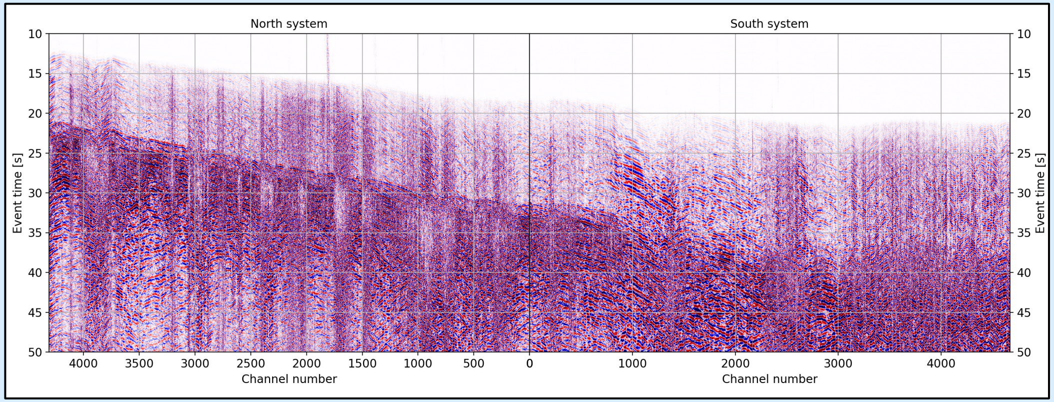
- **100 stations' recordings: ~2.2 MB**
- **10k DAS channels: ~0.22 GB**

How about storing data for the ~20000/yr earthquakes?

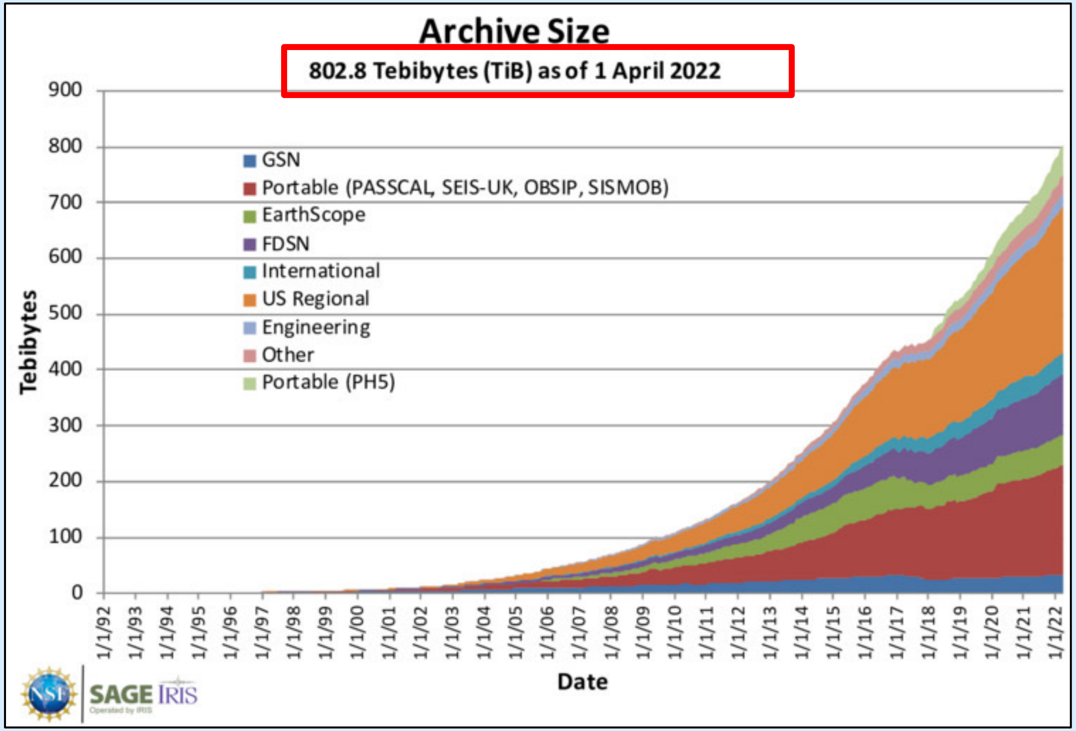
A single 10k-channel DAS would need:

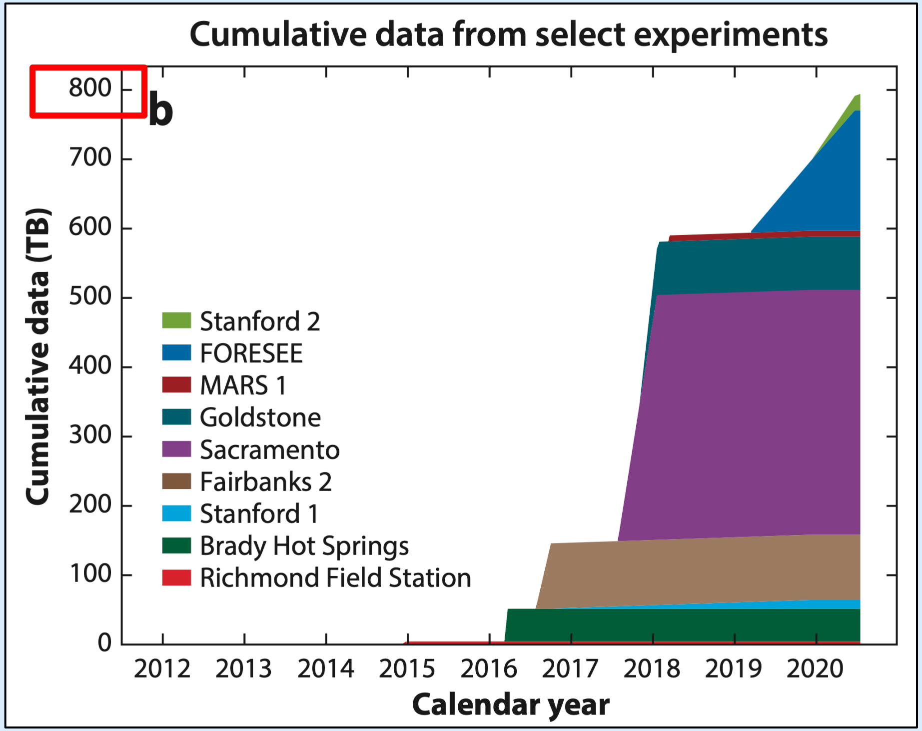
~4.3TB/yr





The IRIS vs DAS database



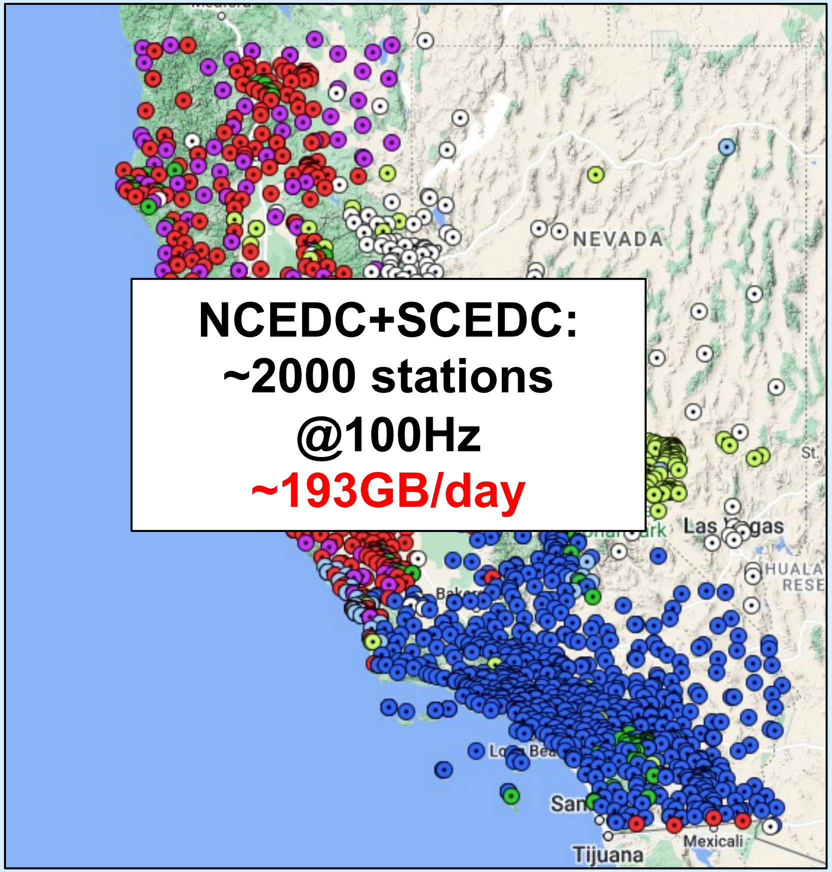


IRIS

Lindsey and Martin, 2021

Conventional networks and DAS

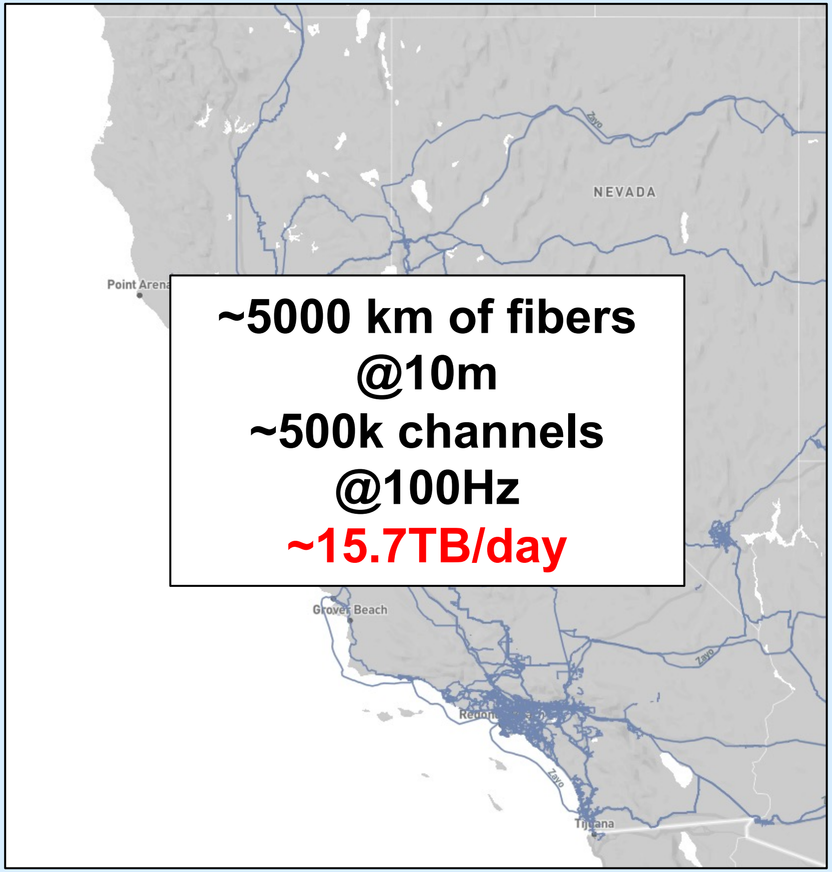
Conventional network



NCEDC+SCEDC:
~2000 stations
@100Hz
~193GB/day

IRIS

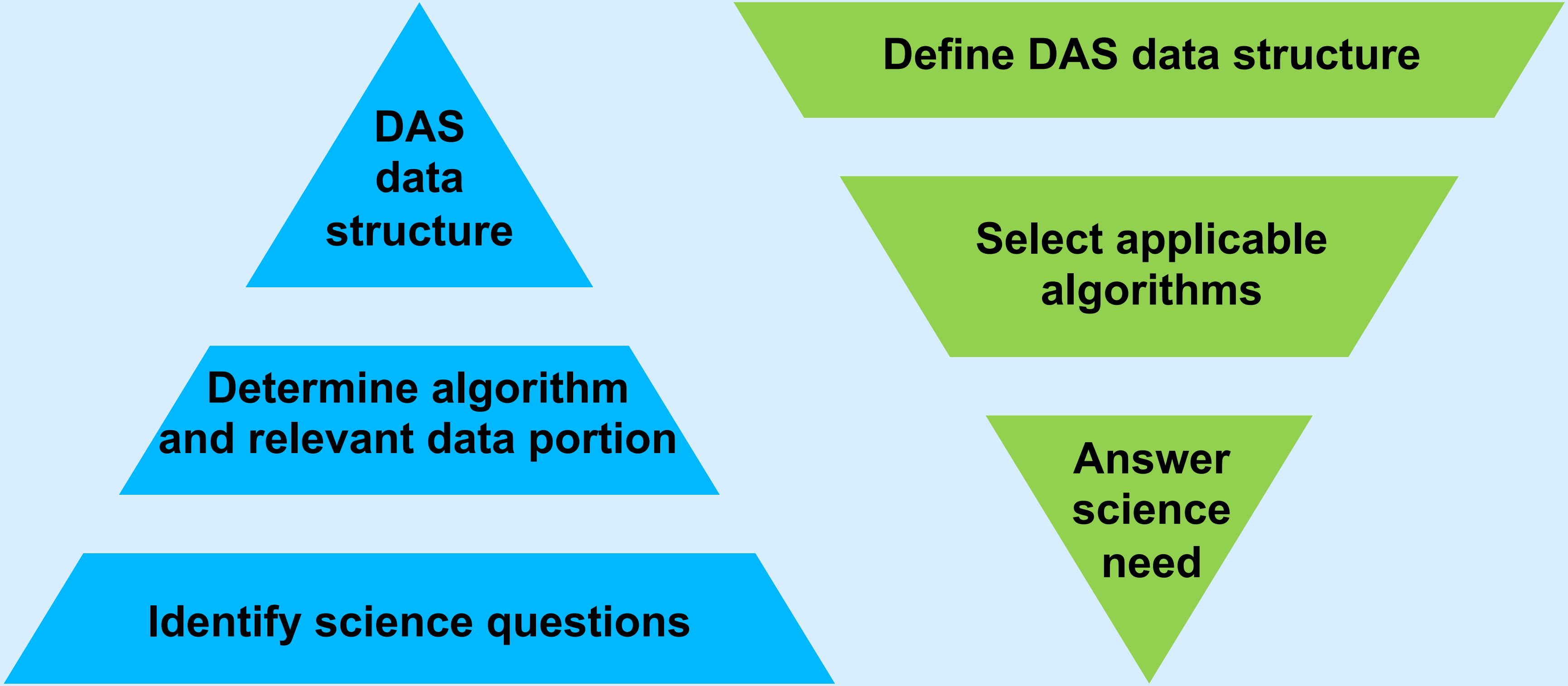
DAS network



~5000 km of fibers
@10m
~500k channels
@100Hz
~15.7TB/day

https://www.infrapedia.com/app

How can we handle DAS data volumes?



Top-down Approach

Define DAS data structure

Select applicable algorithms

Answer science need

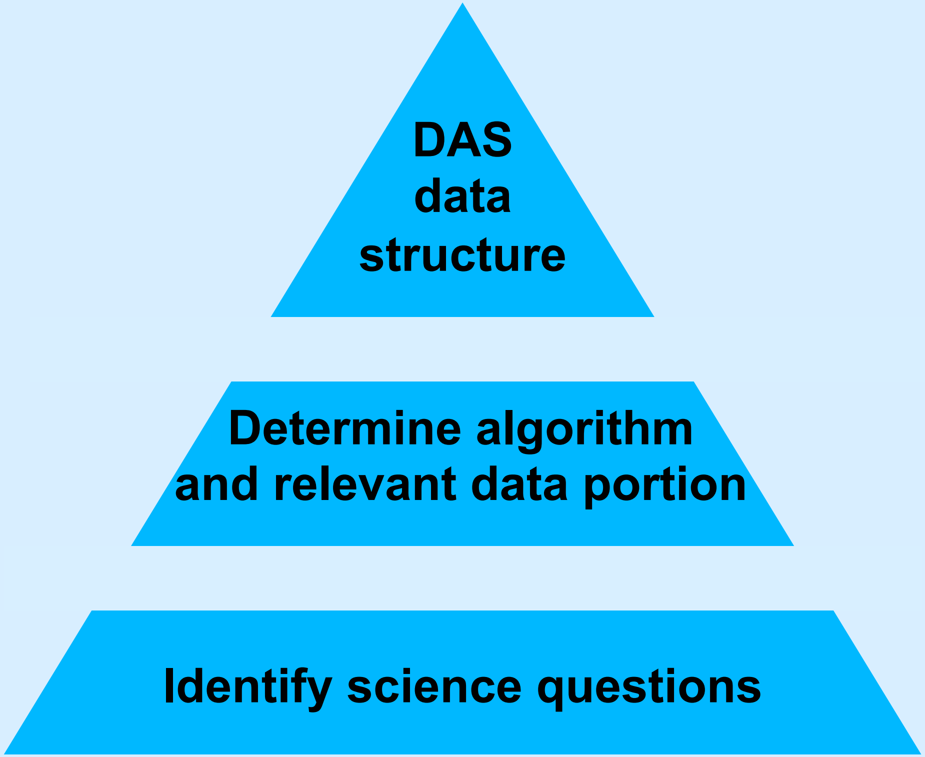
DAS data structure

Determine algorithm and relevant data portion

Identify science questions

Bottom-Up Approach

How can we handle DAS data volumes?



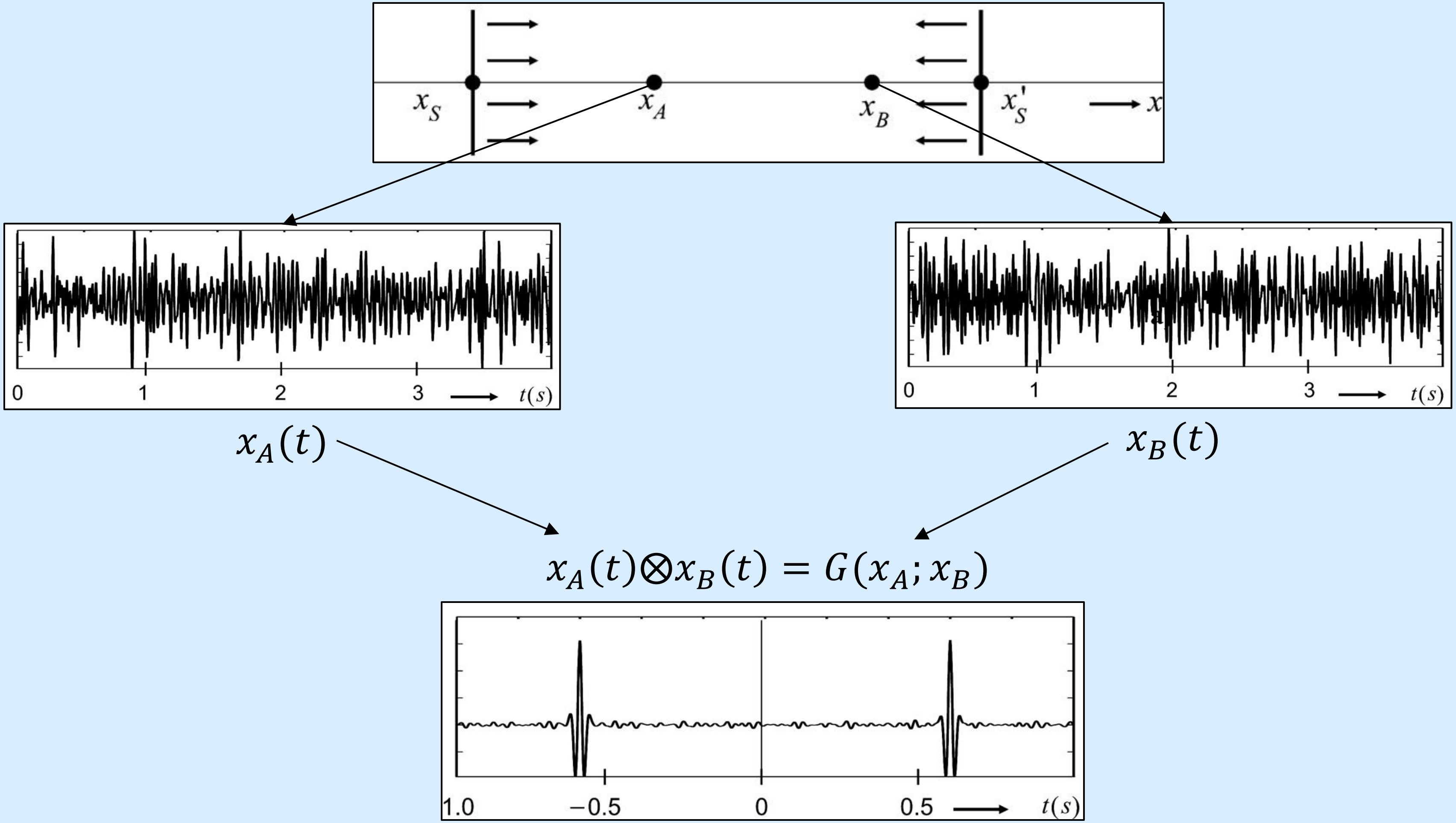
DAS
data
structure

Determine algorithm and relevant data portion

Identify science questions

Bottom-Up Approach

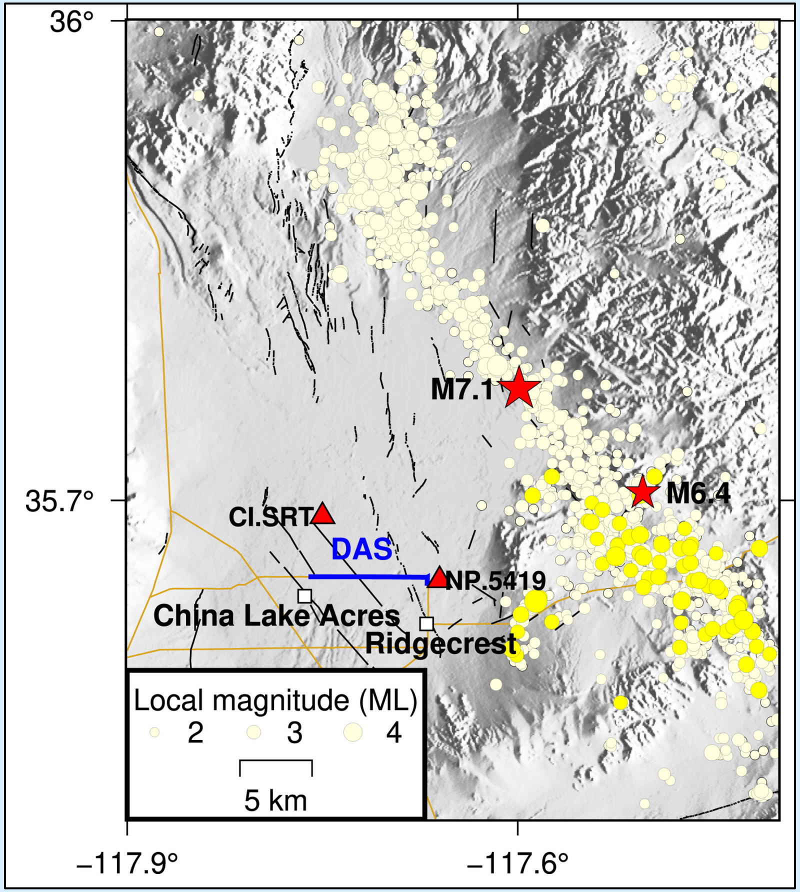
Noise interferometry

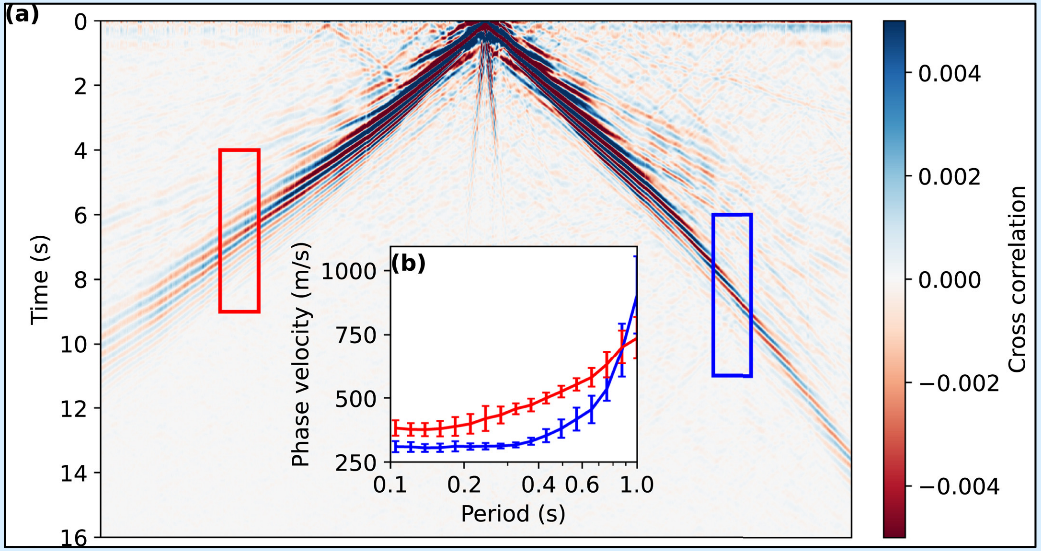


$$x_A(t) \otimes x_B(t) = G(x_A; x_B)$$

Wapenaar et al., 2010

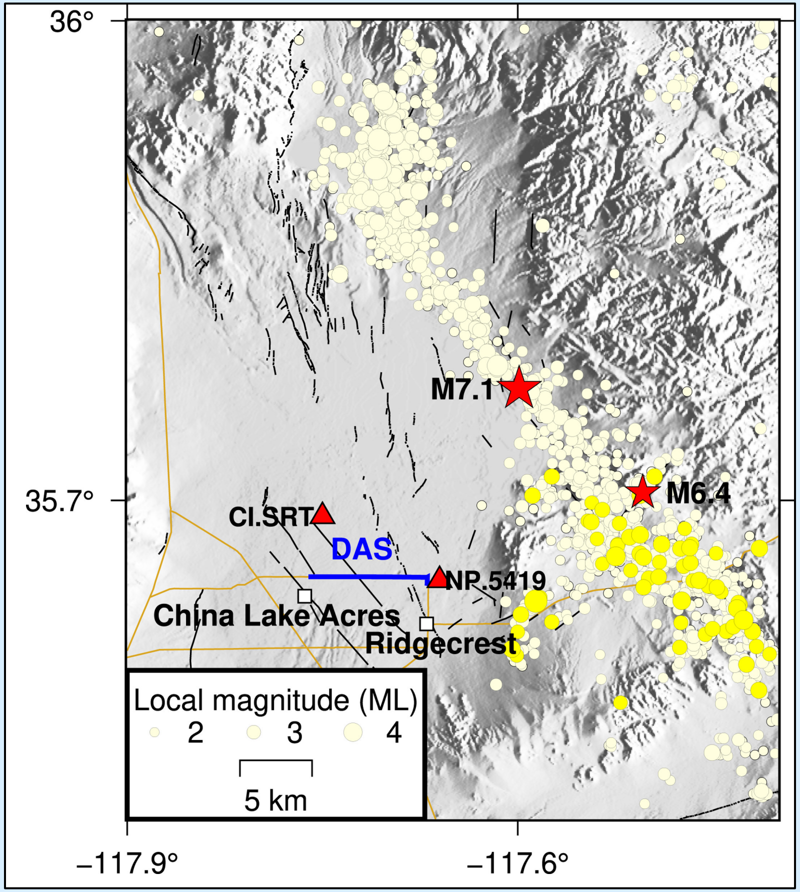
Noise interferometry with DAS

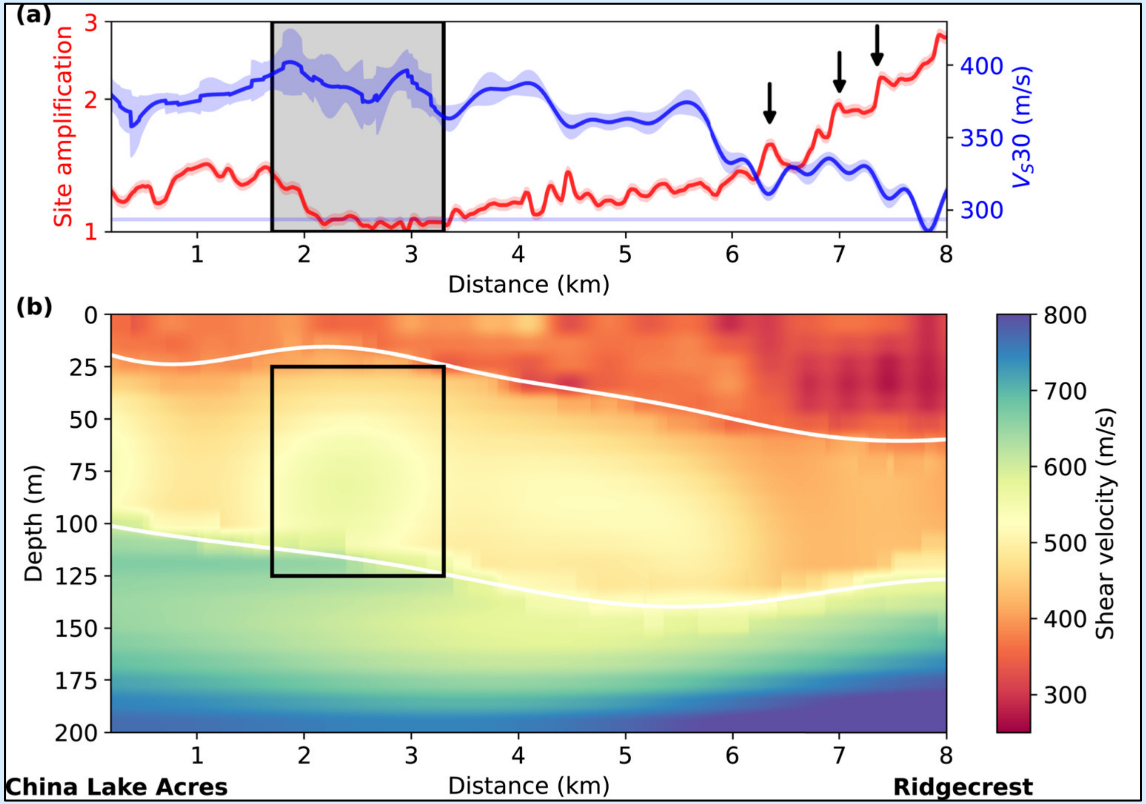




Yang et al., 2021

Noise interferometry with DAS





Yang et al., 2021

Noise interferometry with DAS

What are the challenges of noise interferometry with DAS?

- **The amount of data to be processed**

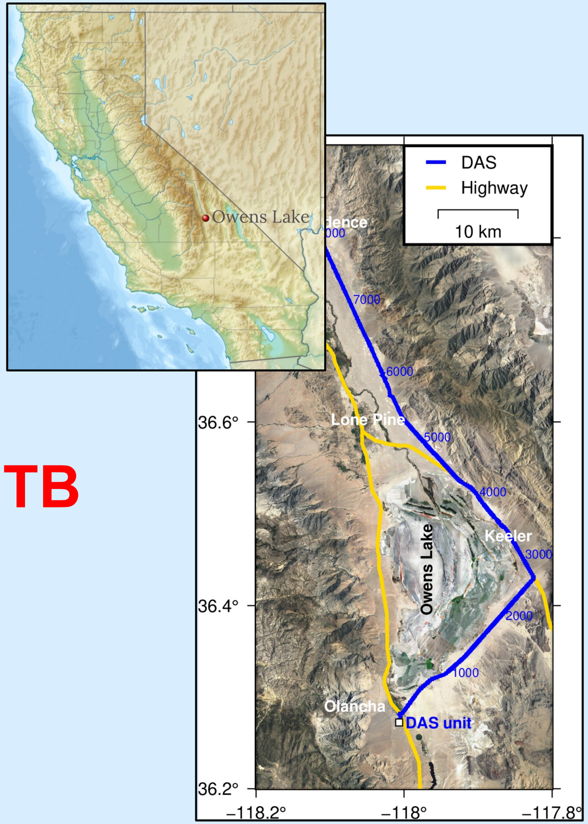
  ~100TB of data

- **The number of channels to cross-correlate**

  10000 channels @ 100Hz => <2TB

- **Storing CCs for time-lapse studies**

  Daily CCs => ~800TB/yr



Yang et al., in prep.

Noise interferometry with DAS

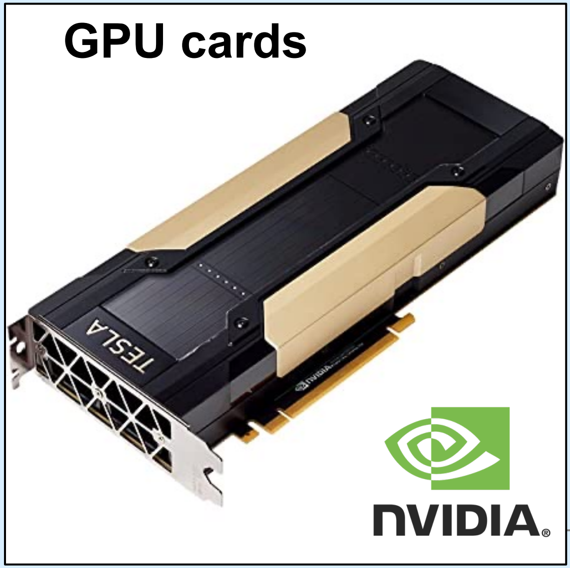
- **The amount of data to be processed**

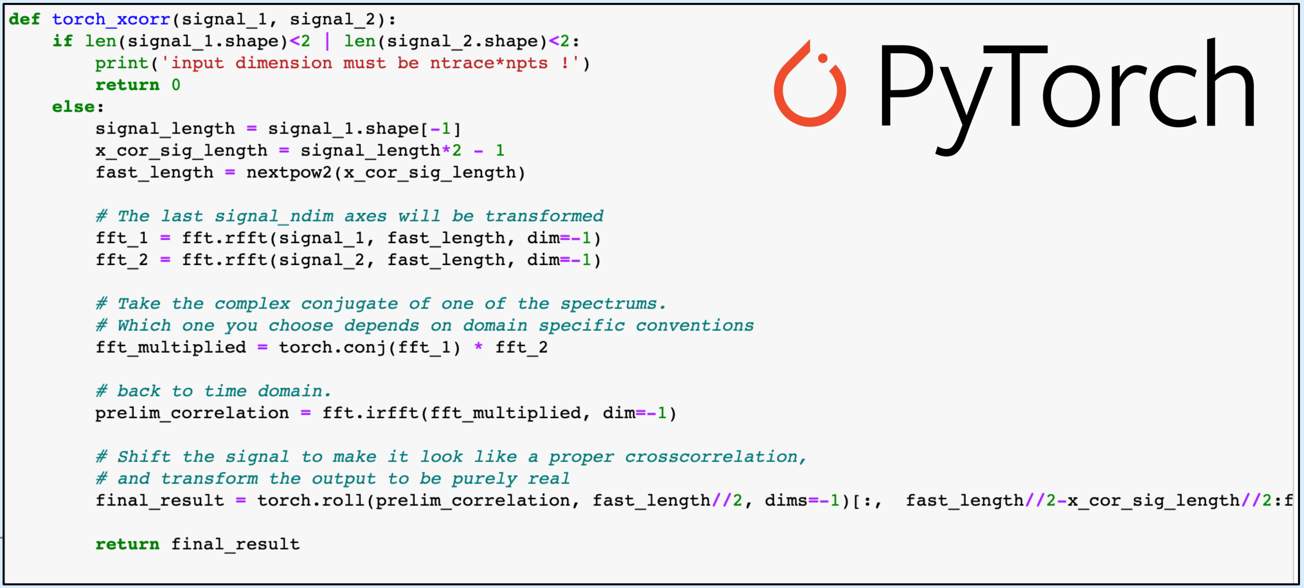
$$x_A(t) \otimes x_B(t) = G(x_A; x_B) = X_A(\omega) * X_B(\omega)$$

Single-frequency 1 month data: 4.5 TFlops

GPU V100: 15.7 TFlops

GPU cards

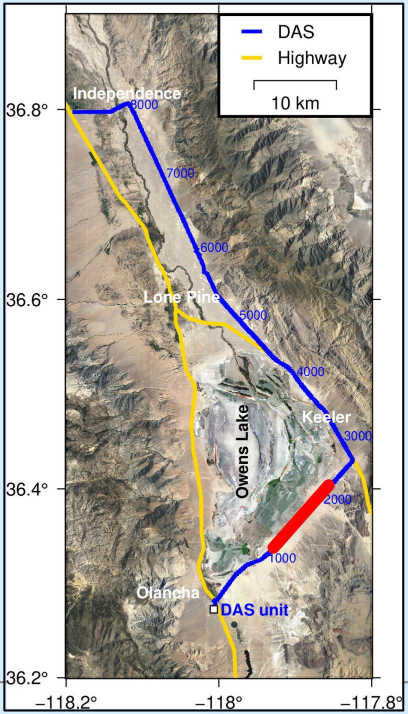


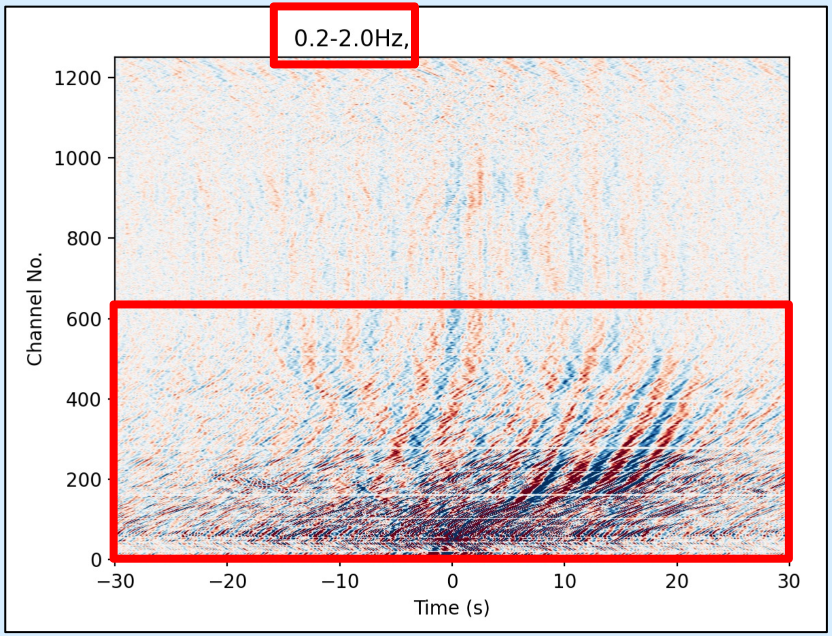


```
def torch_xcorr(signal_1, signal_2):
    if len(signal_1.shape)<2 | len(signal_2.shape)<2:
        print('input dimension must be ntrace*npts !')
        return 0
    else:
        signal_length = signal_1.shape[-1]
        x_cor_sig_length = signal_length*2 - 1
        fast_length = nextpow2(x_cor_sig_length)

        # The last signal_ndim axes will be transformed
        fft_1 = fft.rfft(signal_1, fast_length, dim=-1)
        fft_2 = fft.rfft(signal_2, fast_length, dim=-1)

        # Take the complex conjugate of one of the spectrums.
        # Which one you choose depends on domain specific conventions
        fft_multiplied = torch.conj(fft_1) * fft_2

        # back to time domain.
        prelim_correlation = fft.irfft(fft_multiplied, dim=-1)

        # Shift the signal to make it look like a proper crosscorrelation,
        # and transform the output to be purely real
        final_result = torch.roll(prelim_correlation, fast_length//2, dims=-1)[:,  fast_length//2-x_cor_sig_length//2:f

        return final_result
```

Noise interferometry with DAS

- **The number of channels to cross-correlate**
  - **Process subarrays**
- **Spatial desampling for seismic scales: from 10m to 200m!**





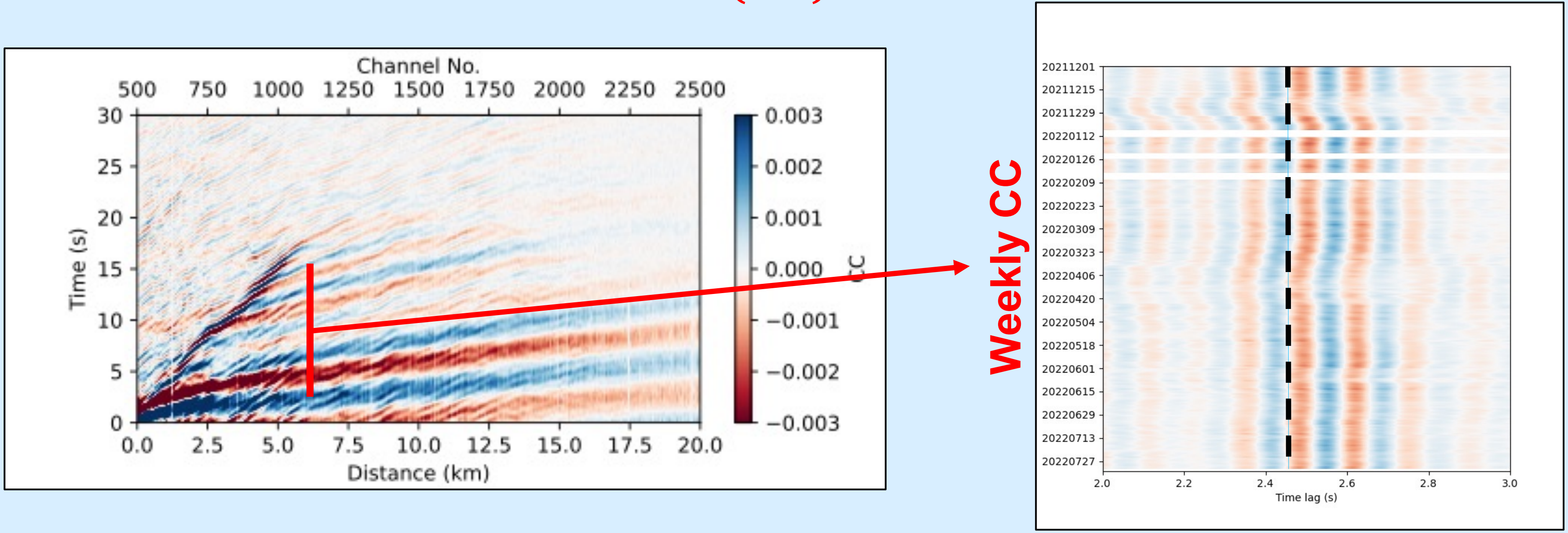
Noise interferometry with DAS

- **Storing CCs for time-lapse studies**

Common-offset channel pairs for entire array

Scale of interest: ~50-200m

$$\mathcal{O}(N^2) \Rightarrow \mathcal{O}(N)$$



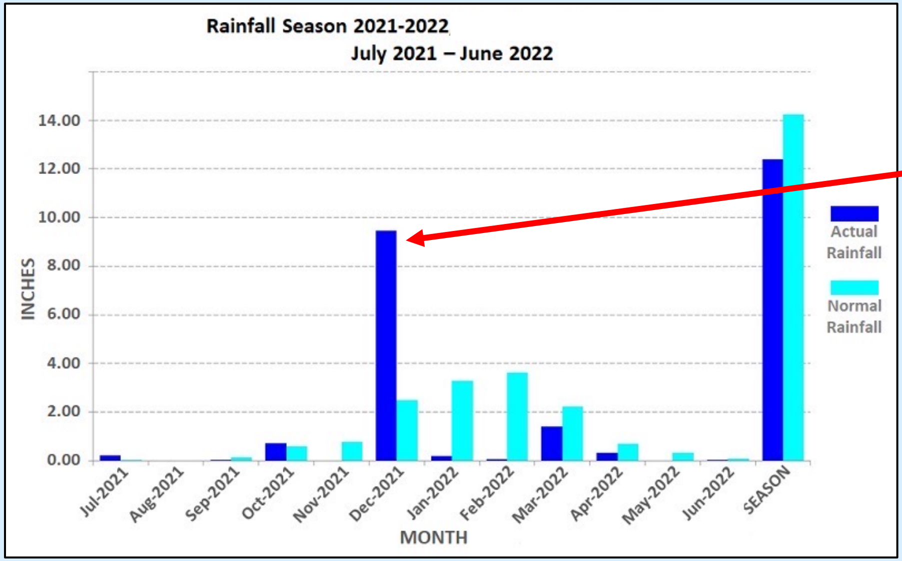
Noise interferometry with DAS

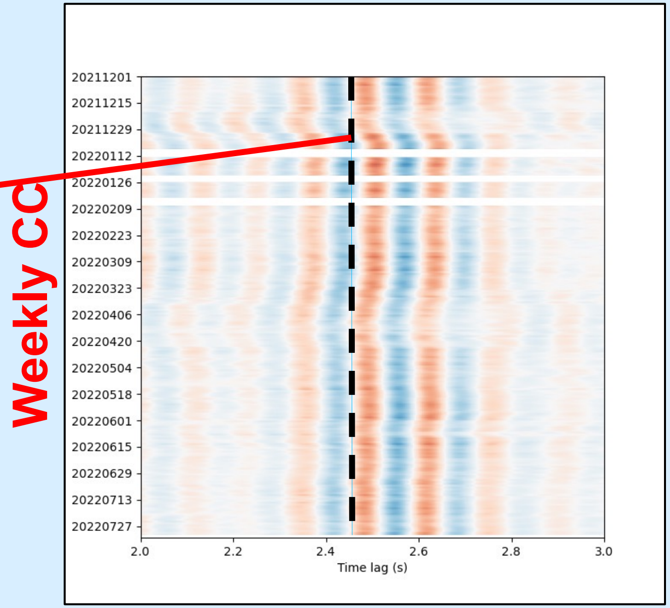
- **Storing CCs for time-lapse studies**

Common-offset channel pairs for entire array

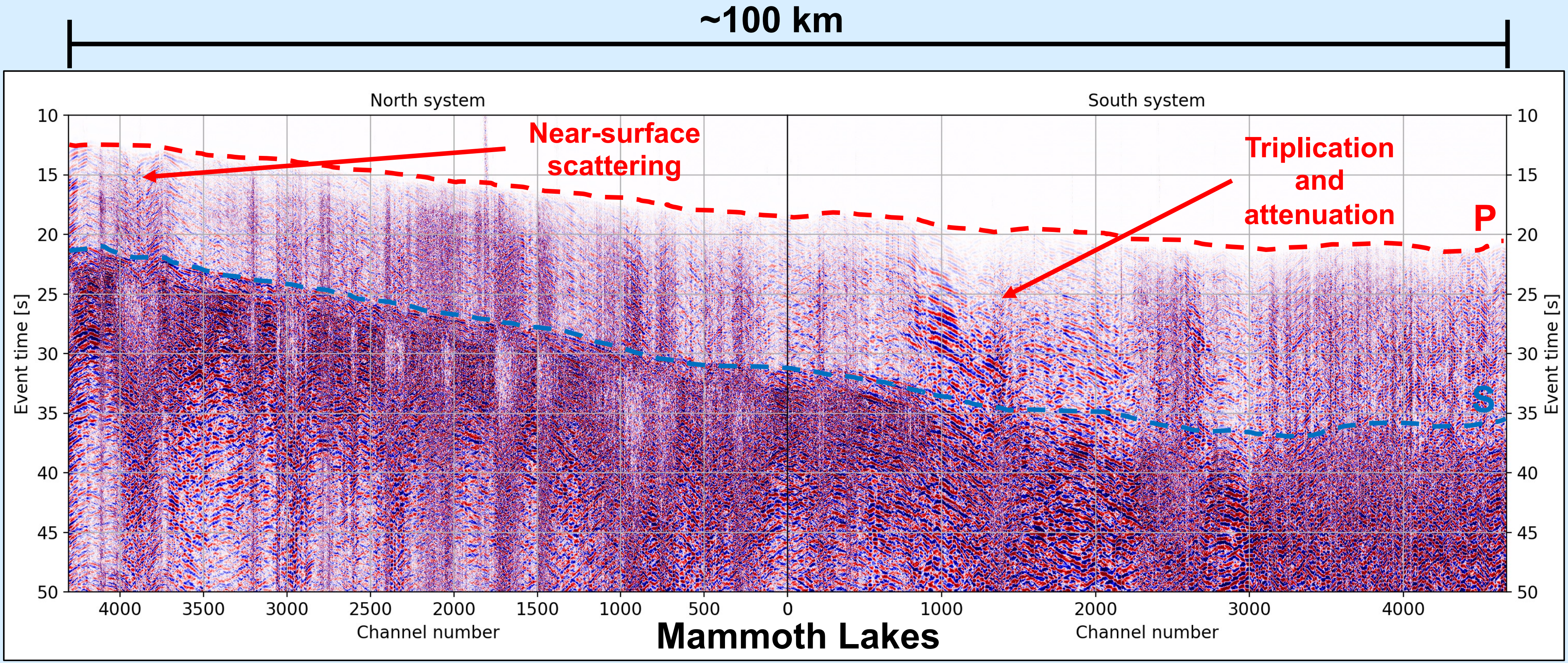
Scale of interest: ~50-200m

$\mathcal{O}(N^2)$ => $\mathcal{O}(N)$



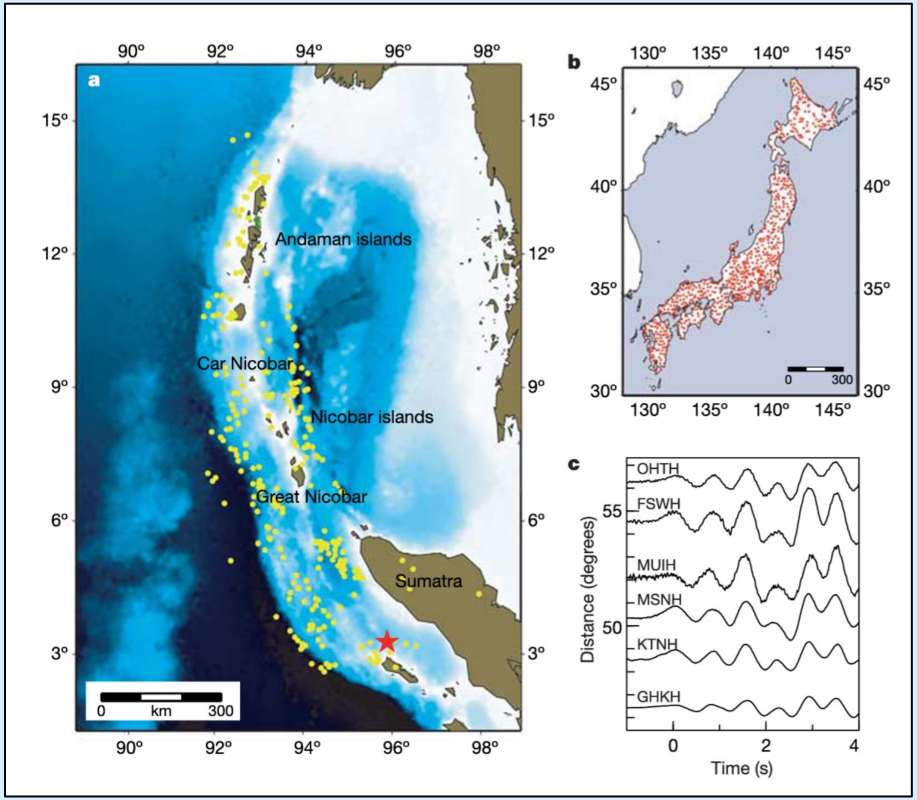


Earthquake DAS data



Back-projection imaging

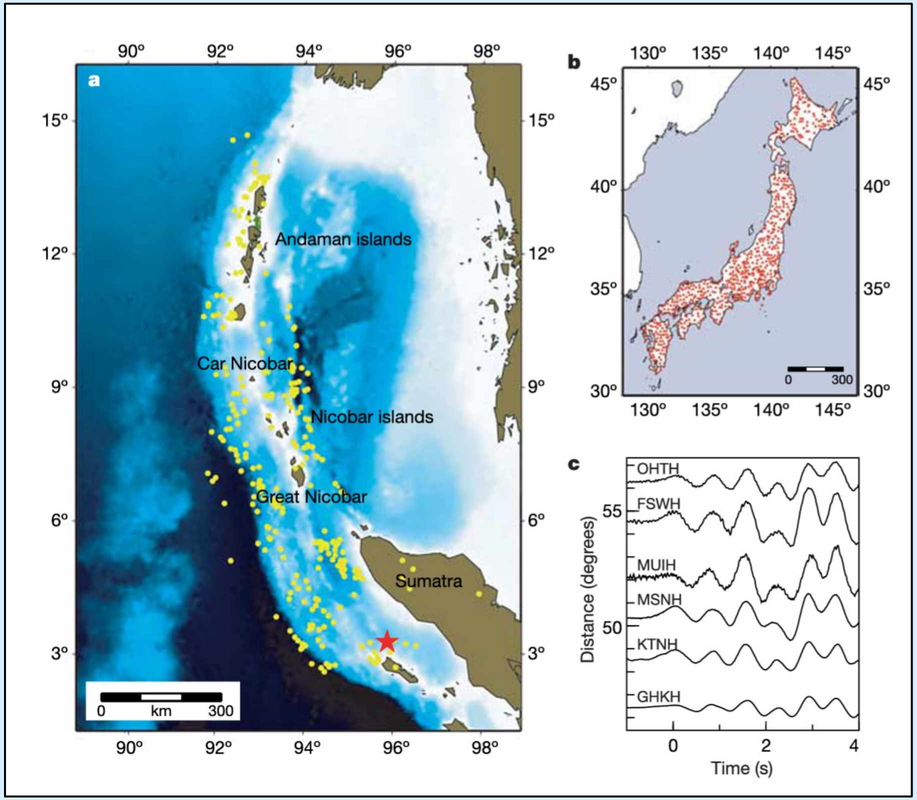
Conventionally, teleseismic waves are used for 2D back-projection imaging



Ishii et al., 2005

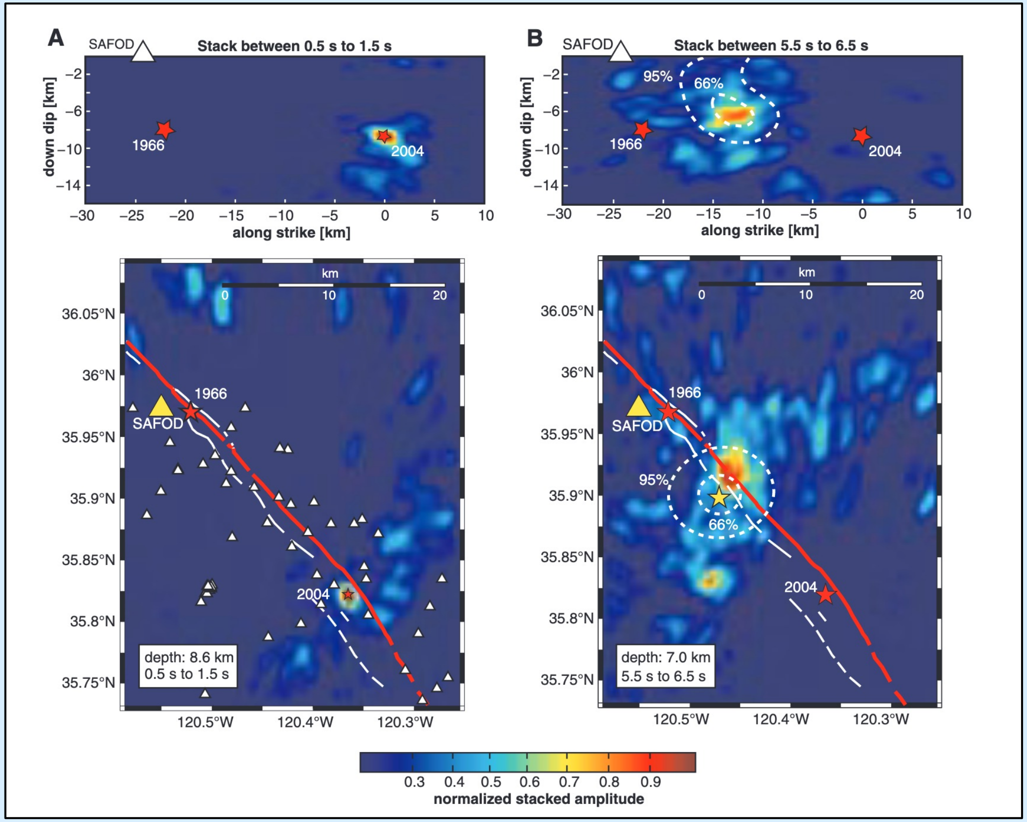
Back-projection imaging

Conventionally, teleseismic waves are used for 2D back-projection imaging



Ishii et al., 2005

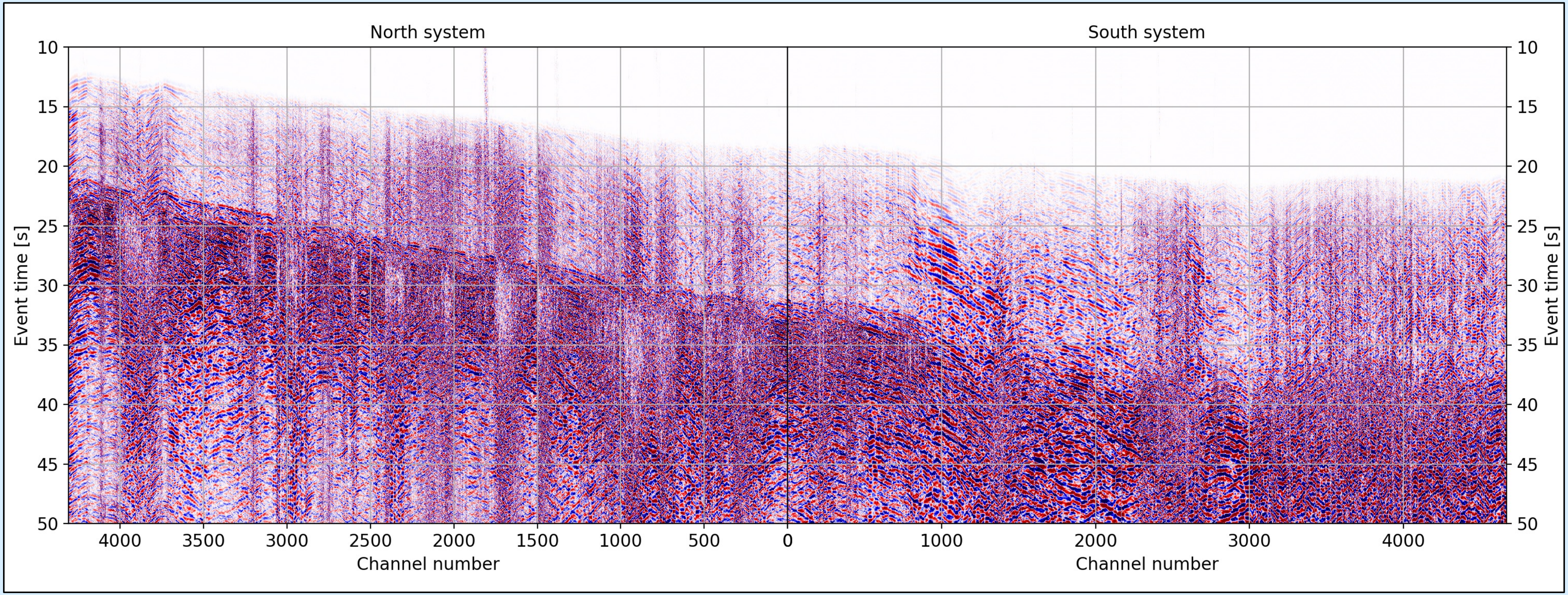
Designed dense arrays allow 3D back-projection of high-frequency energy



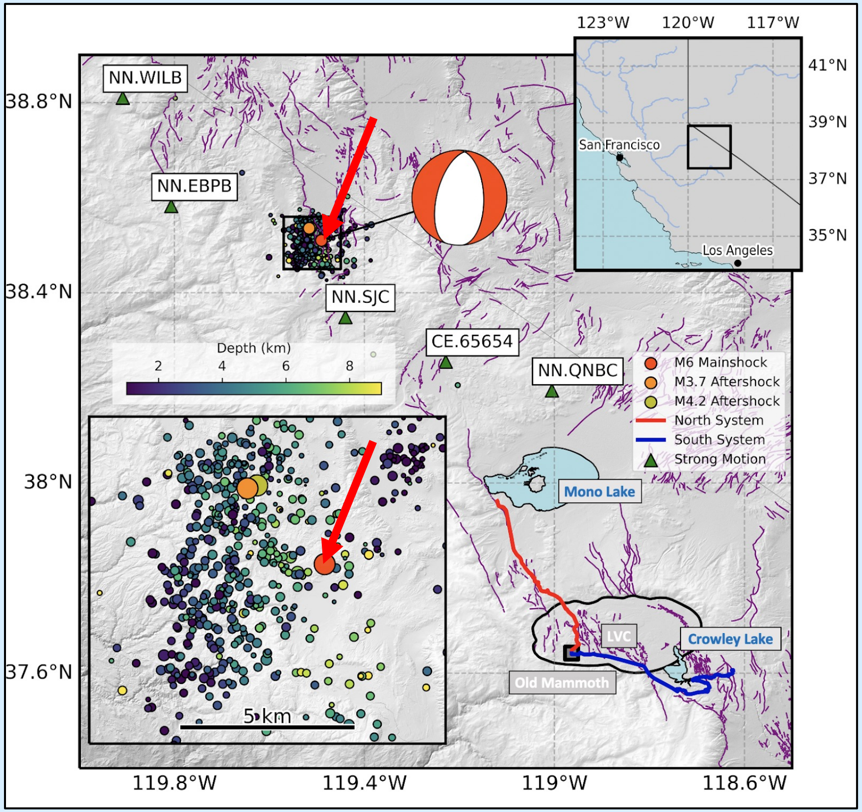
Allmann and Shearer, 2007

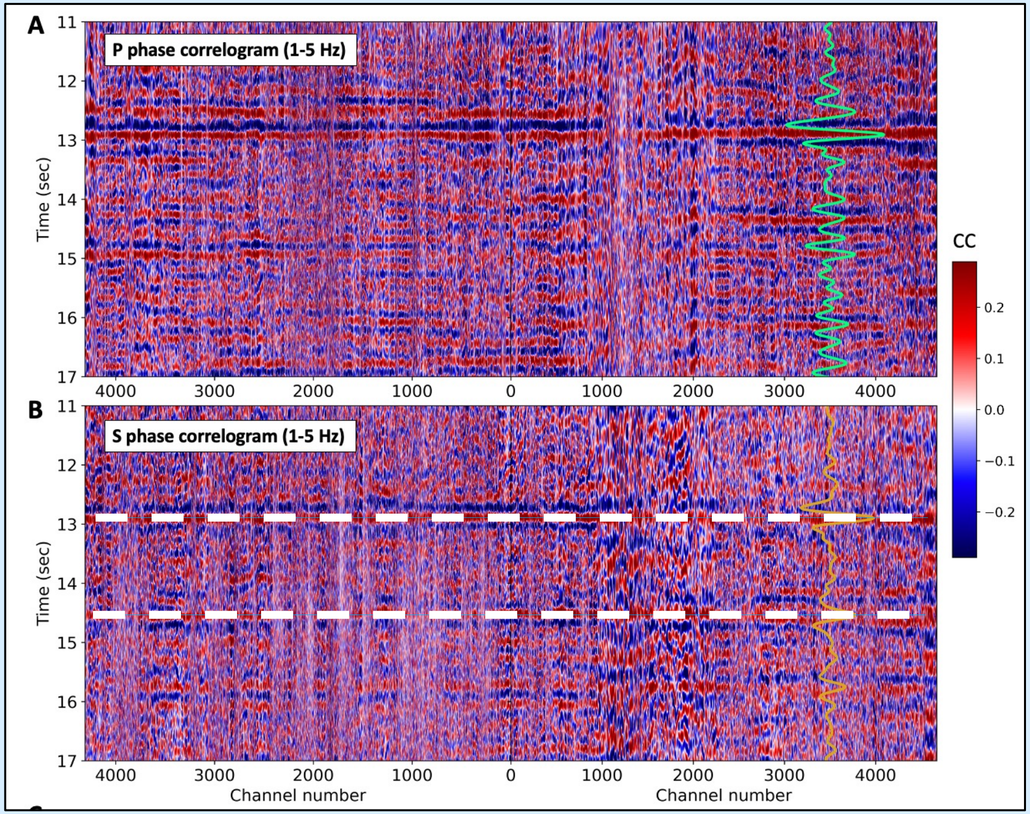
Back-projection imaging with DAS

Would DAS help image high-frequency energy?



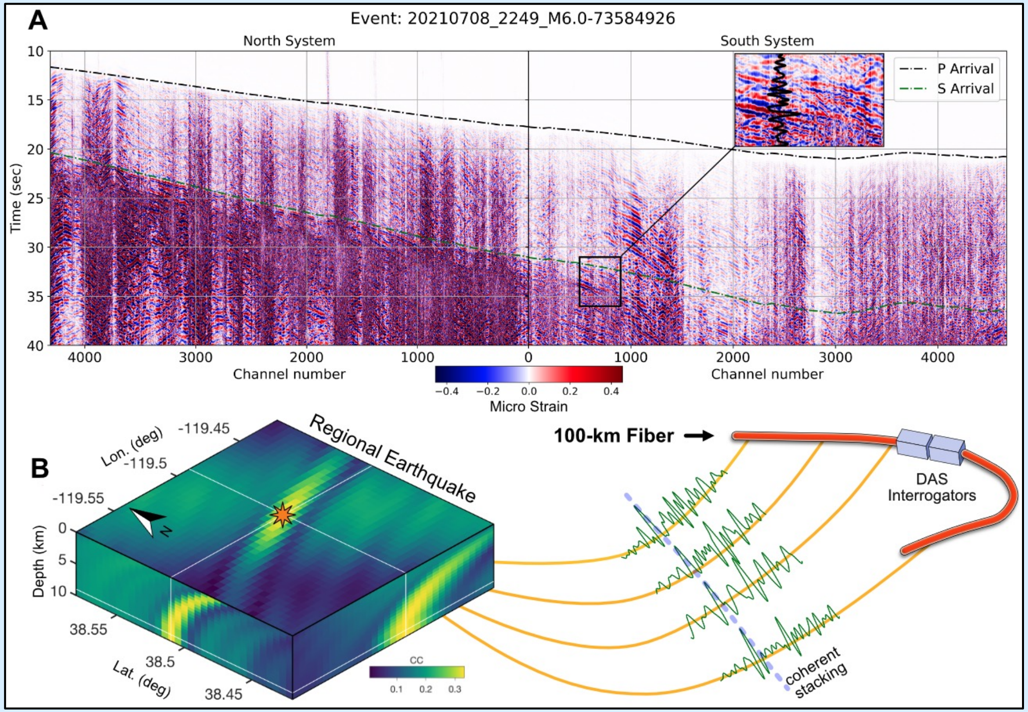
Back-projection imaging with DAS





Li et al., 2022

Back-projection imaging with DAS

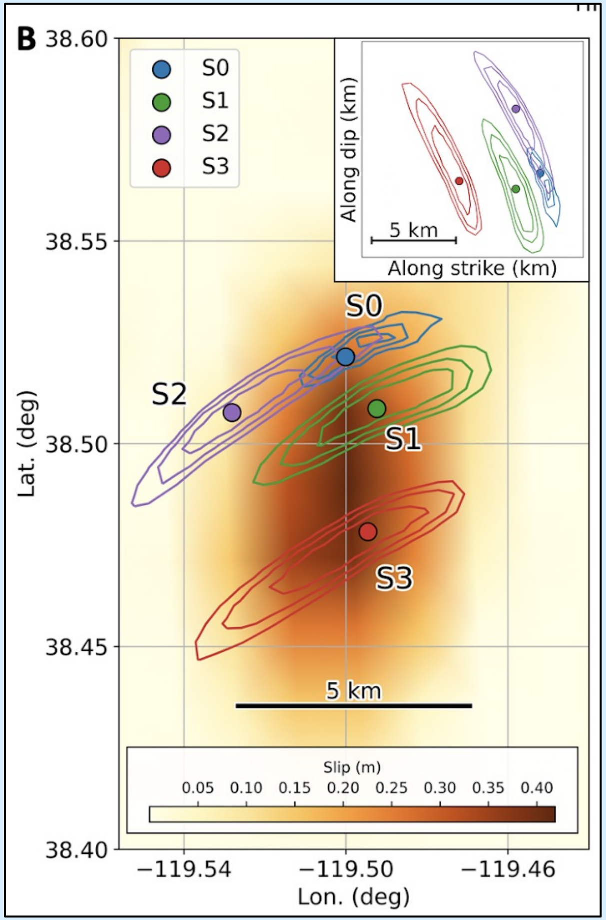


What are the main challenges?

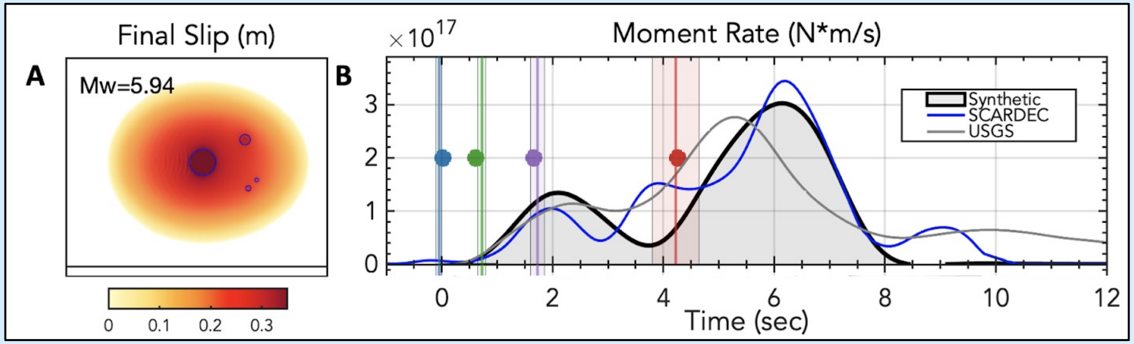
- **Traveltime computation on 3D volume ⇒ $10^6 - 10^7\ N_v$**
- **Conventional station $N_d \approx 10^2$**
  DAS $N_d \approx 10^4$
- **Computational complexity $\mathcal{O}(N_v * N_d)$**
- **Each grid point independent**
  Using GPUs: $\mathcal{O}(N_d)$

Li et al., 2022

Backprojection imaging with DAS

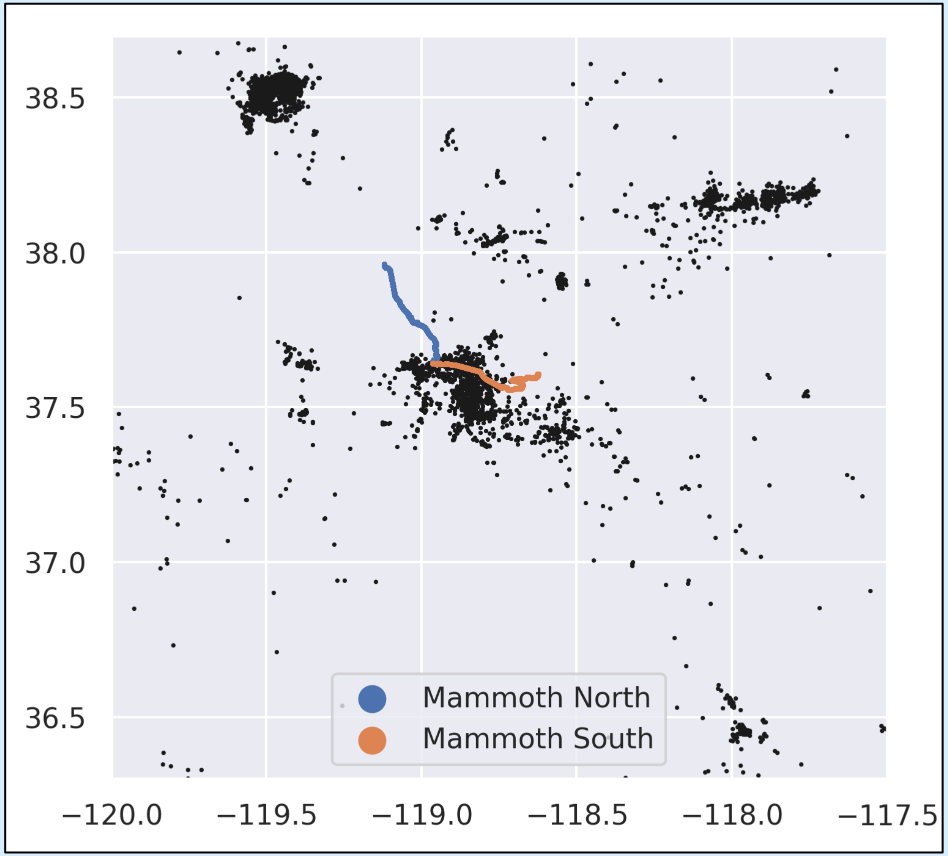


Imaging the high-frequency rupture process!



Li et al., 2022

How do we use DAS event data?



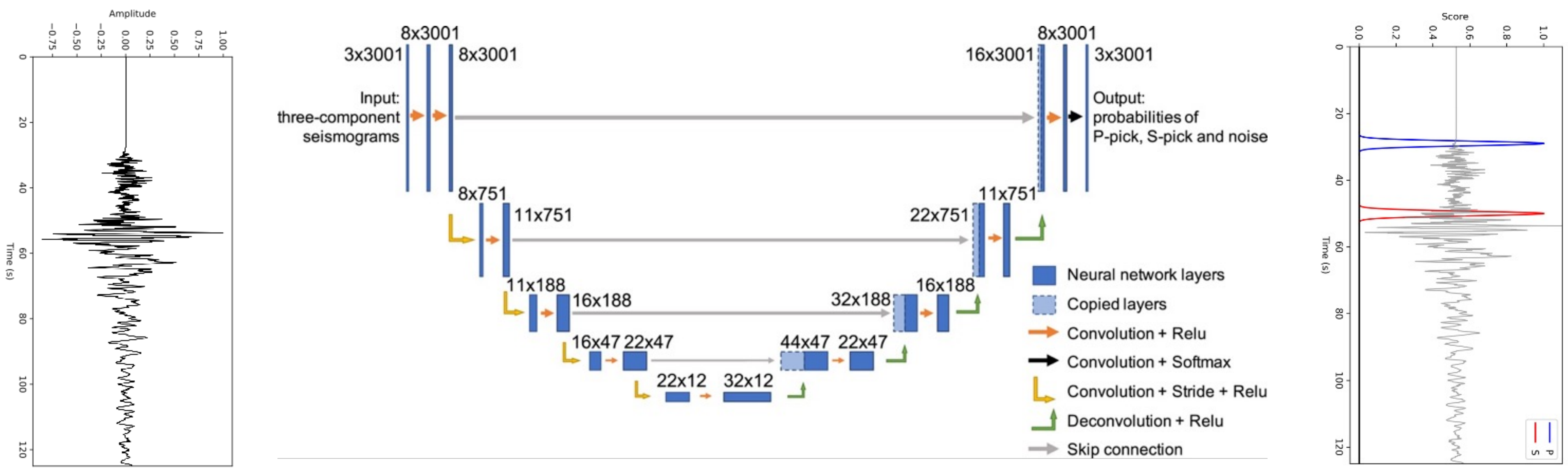
~8000 events
P- and S-wave picks:

Stations => 1.6 million

DAS => 160 millions

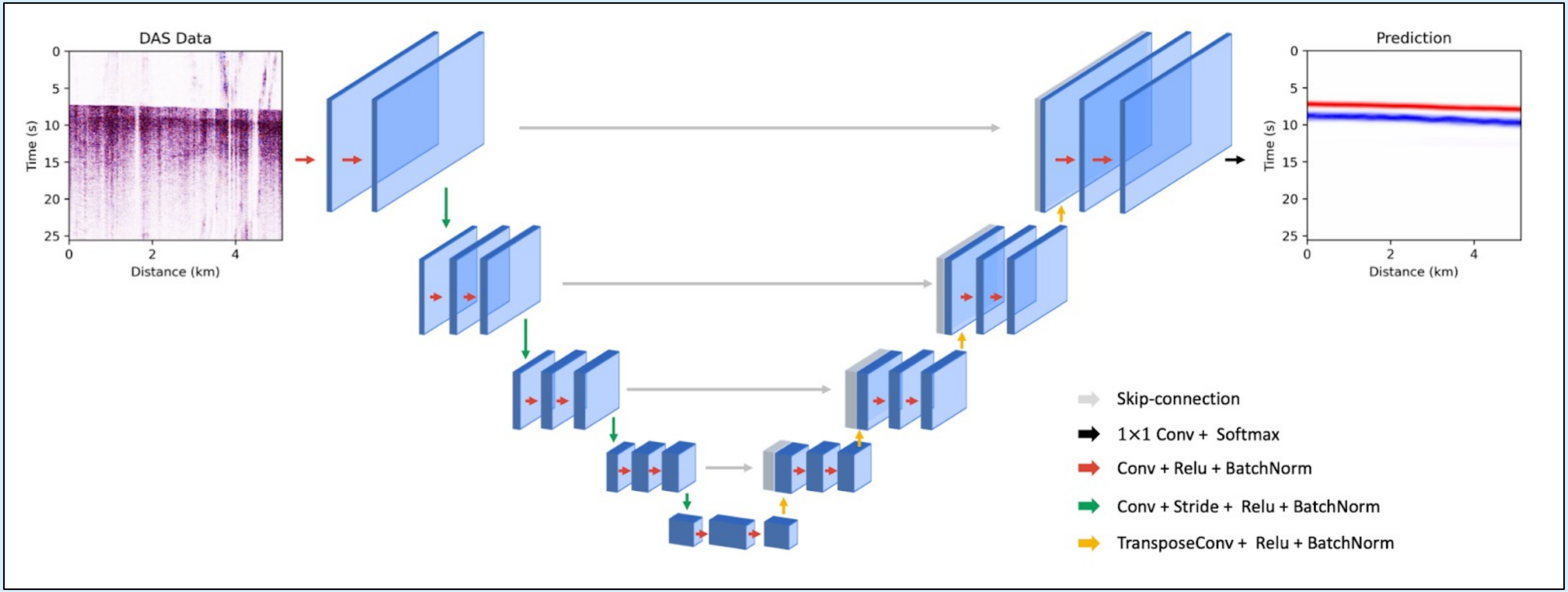
A DAS picking algorithm does not exist!

Accurate event picking: PhaseNet



Zhu and Beroza, 2018

Accurate event picking: PhaseNet-DAS

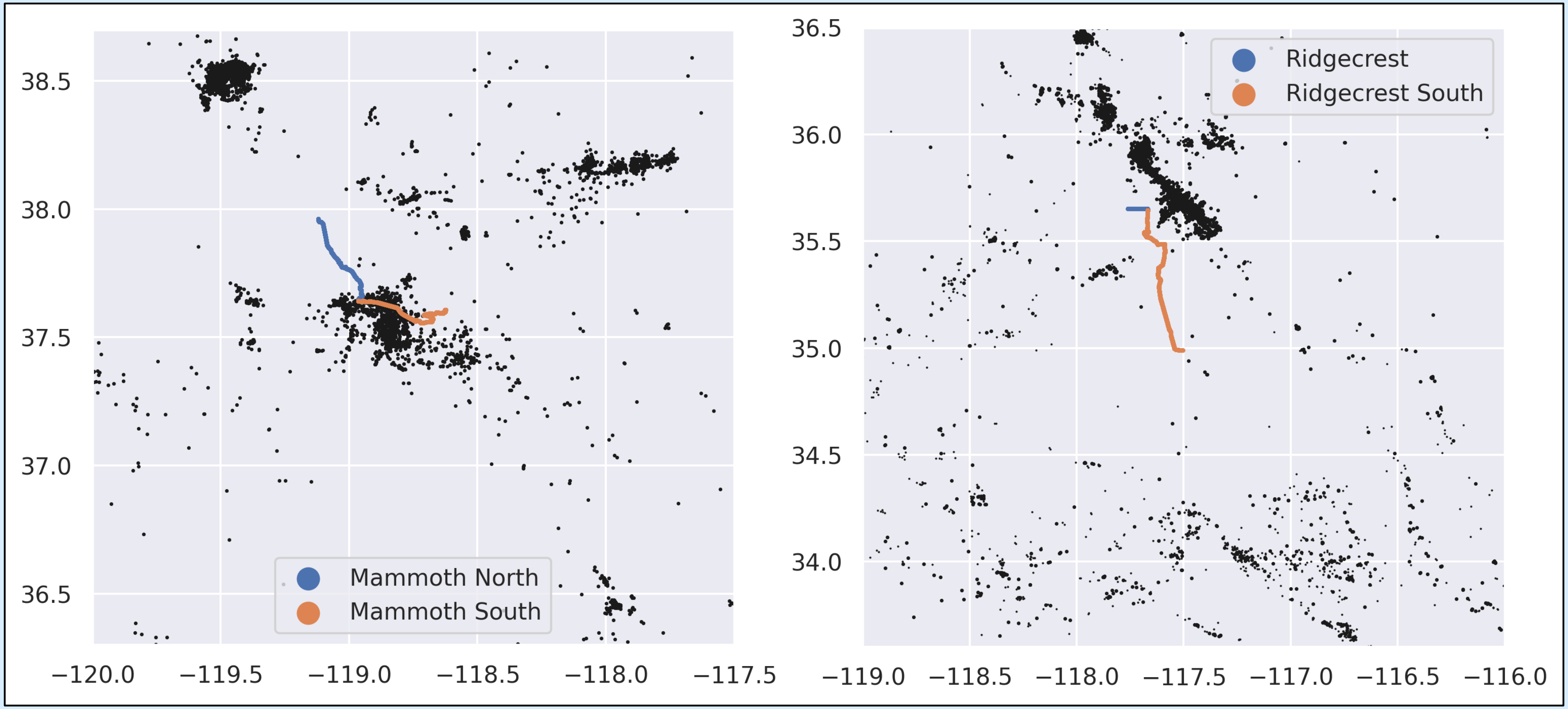


Zhu et al., in prep.

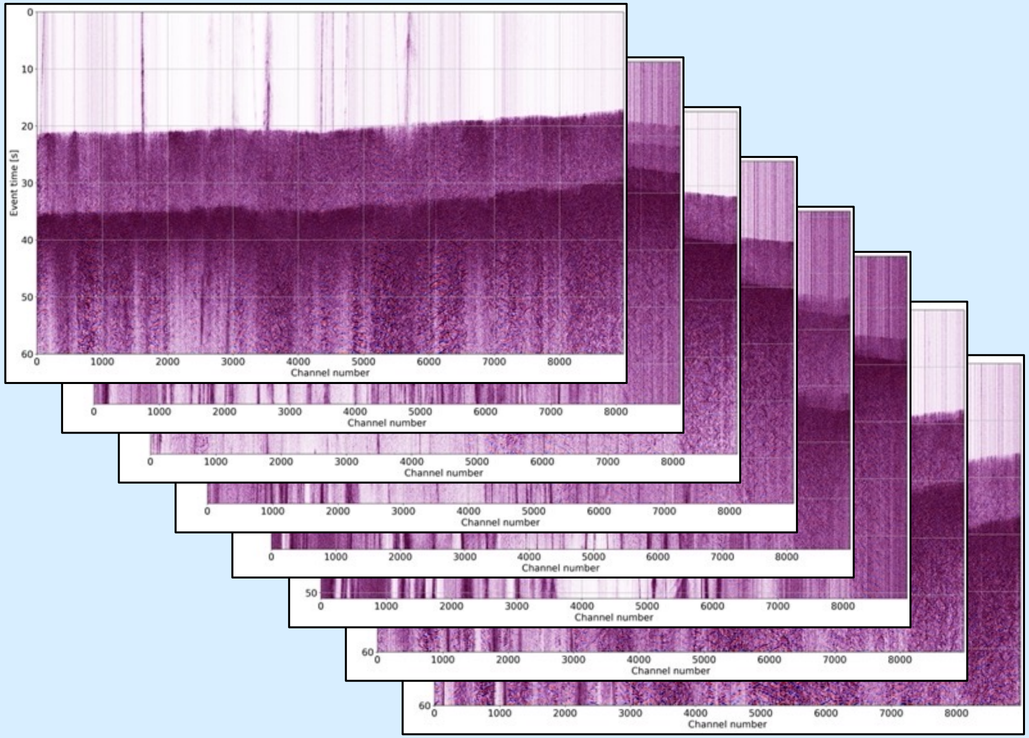
Accurate event picking: PhaseNet-DAS

In the last 1.5 years, we recorded more than 21000 earthquakes!

@100Hz => 4.2TB



Accurate event picking: PhaseNet-DAS



Incredible dataset but challenging to tackle computationally!

4.2TB of data

PyTorch

NVIDIA

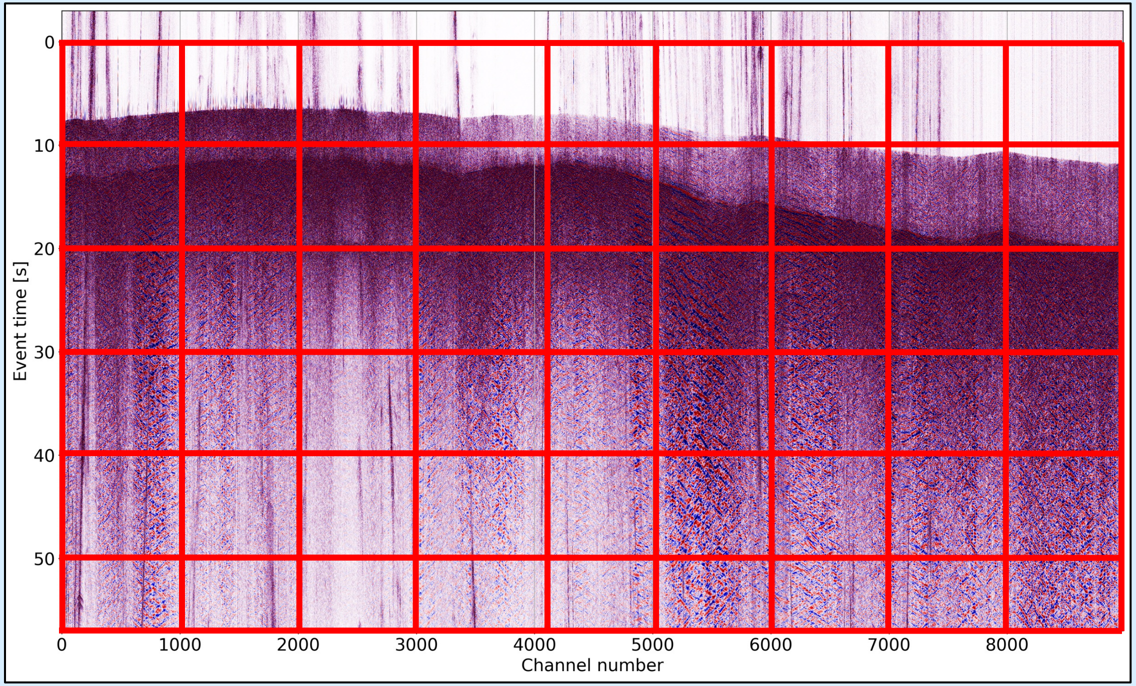
GPU memory ~ 16-32 GB
Single event: ~250 MB

Very few training examples can be stored!

Zhu et al., in prep.

Accurate event picking: PhaseNet-DAS

Employing a patching approach
Faster training and fewer model parameters!



Incredible dataset but challenging to tackle computationally!

4.2TB of data

PyTorch

NVIDIA

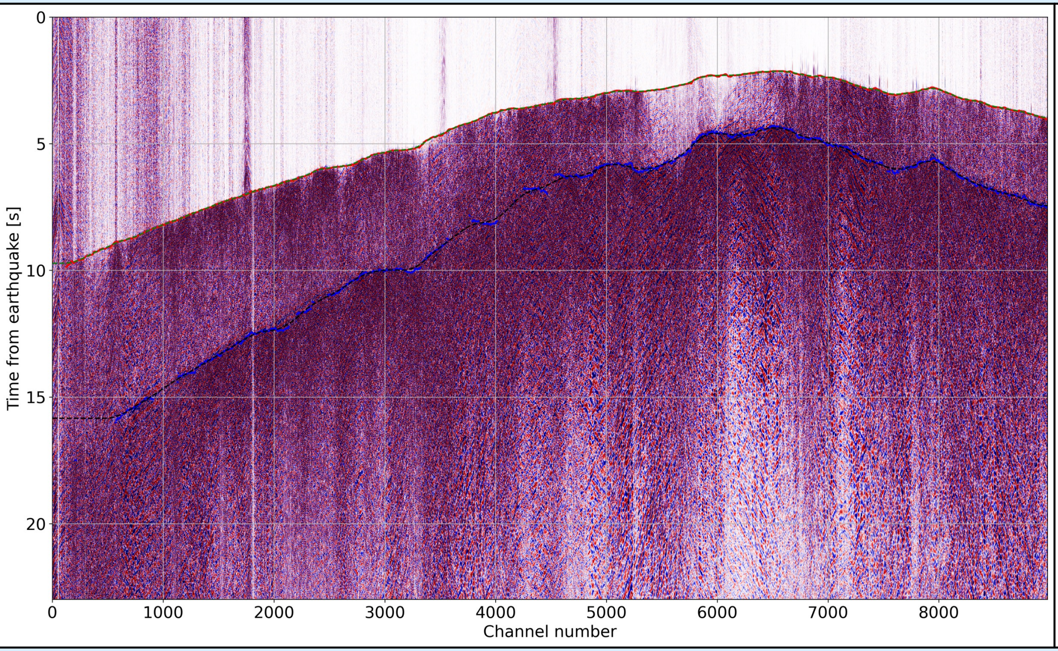
GPU memory ~ 16-32 GB
Single event: ~250 MB

Very few training examples can be stored!

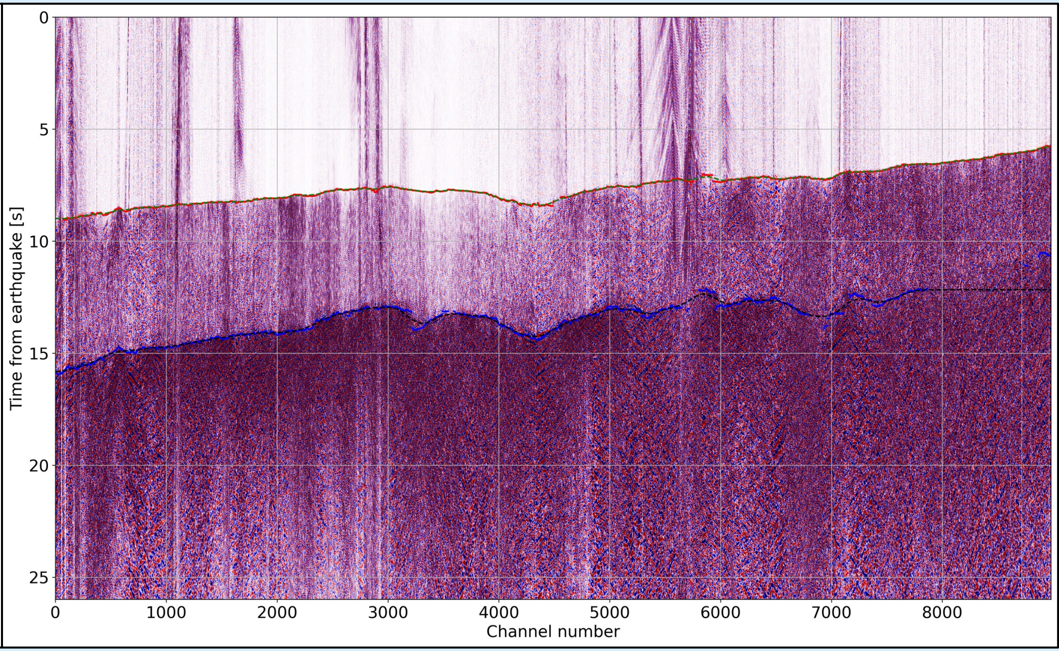
Zhu et al., in prep.

Accurate event picking: PhaseNet-DAS

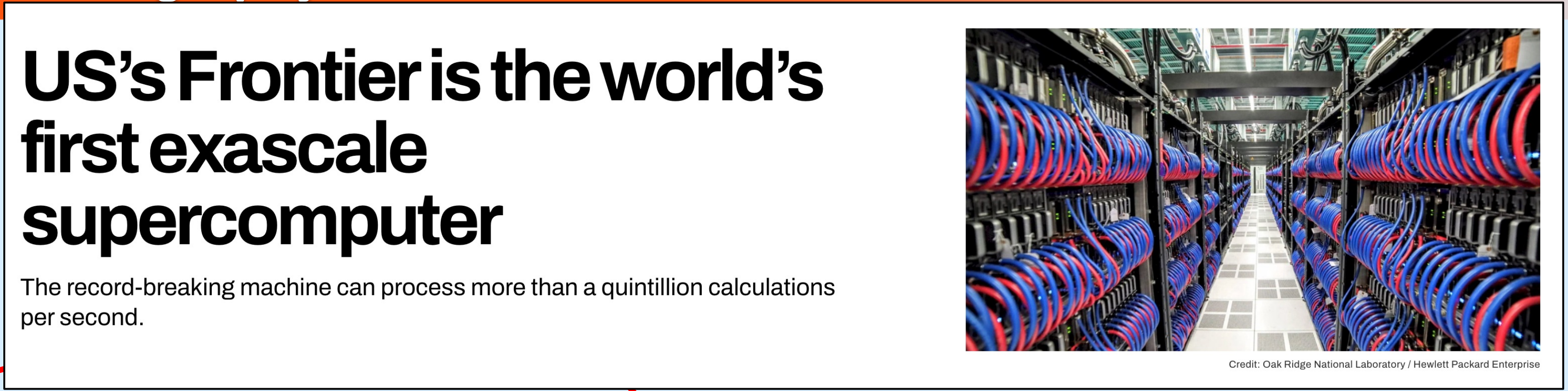
Local event



Regional event



3D tomography with DAS



US's Frontier is the world's first exascale supercomputer

The record-breaking machine can process more than a quintillion calculations per second.

Credit: Oak Ridge National Laboratory / Hewlett Packard Enterprise

$$\mathrm{Q_{DD}} = \begin{bmatrix} 1 & -1 & \bullet & \bullet & \bullet & 0 \\ 1 & \bullet & \bullet & -1 & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & \bullet & 1 & \bullet & \bullet & -1 \end{bmatrix}$$

Common size $N \approx 10^5$

Size with DAS $N \approx 10^9$

Computational complexity $O(N^3)$

Lower-bound runtime: ~31 years

Zhang and Thurber, 2006

3D tomography with DAS

What about using TomoDD with this dataset?

$$\boxed{\mathbf{Q}_{\mathbf{DD}}\mathbf{A}\Delta\mathbf{X} + \mathbf{Q}_{\mathbf{DD}}\mathbf{C}\Delta\mathbf{M} = \mathbf{Q}_{\mathbf{DD}}\Delta\mathbf{T}.}$$

$$\mathbf{Q}_{\mathbf{DD}} = \begin{bmatrix} 1 & -1 & \bullet & \bullet & \bullet & 0 \\ 1 & \bullet & \bullet & -1 & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & \bullet & 1 & \bullet & \bullet & -1 \end{bmatrix}$$

Common size $N \approx 10^5$

Size with DAS $N \approx 10^9$

Computational complexity $\mathcal{O}(N^3)$

Zhang and Thurber, 2006

3D tomography with DAS

Eikonal equation:

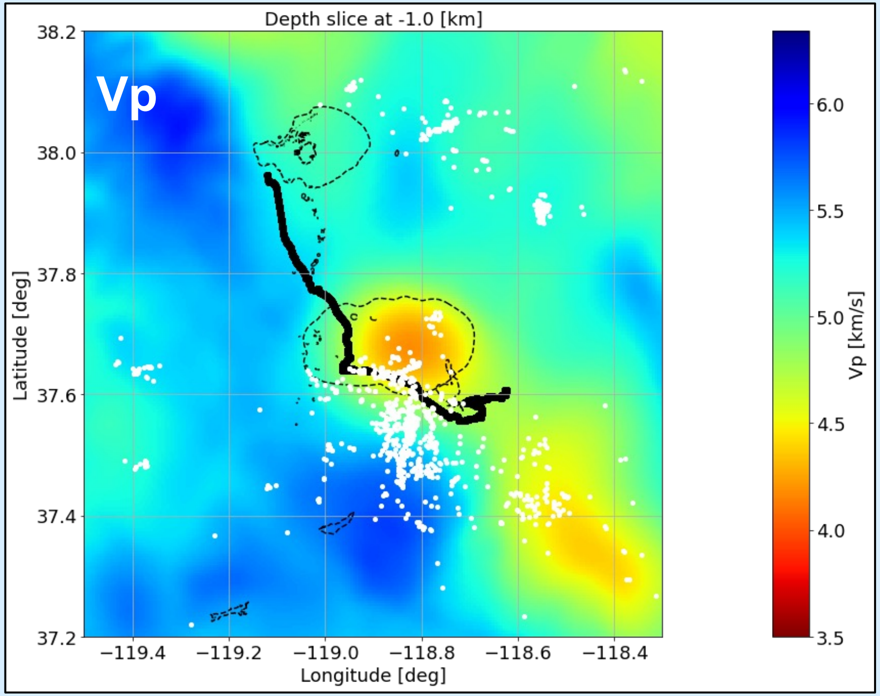
Matrix-free iterative inversion strategy:

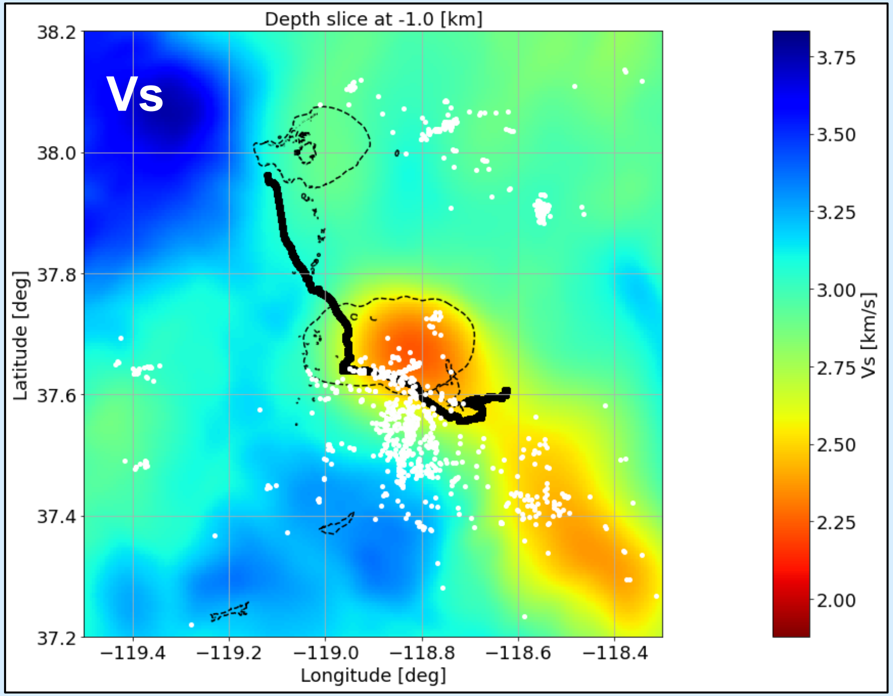
Scalable to HPC clusters!

Biondi et al., in prep

The Long Valley caldera: DAS tomography

Initial guess from CVM

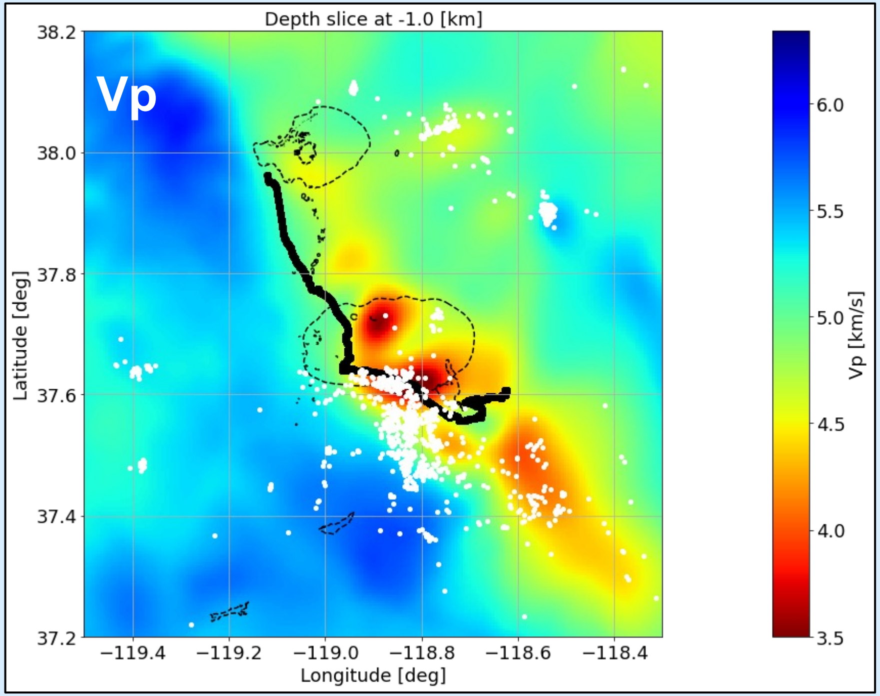


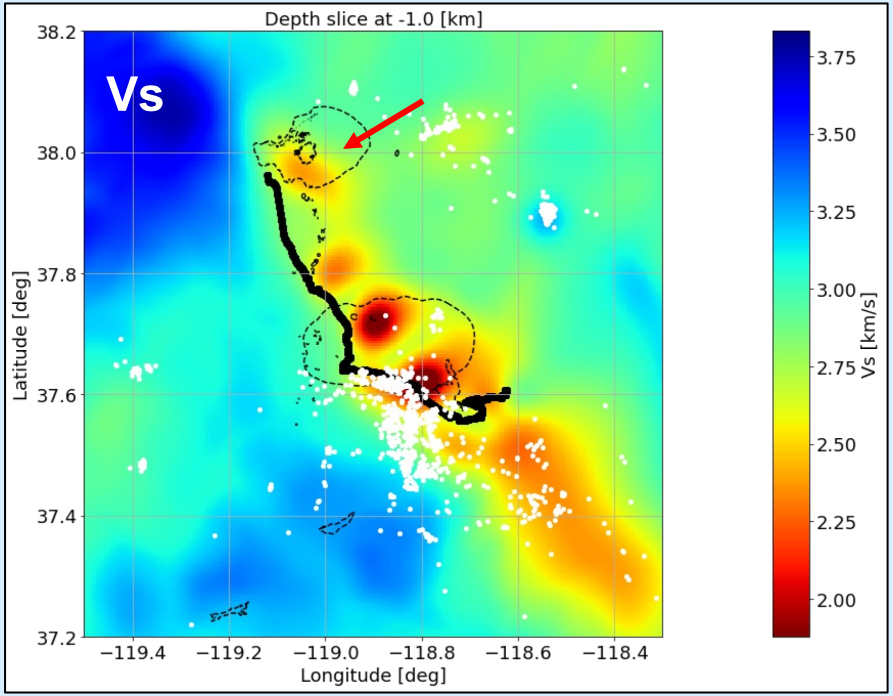


Lee et al., 2014

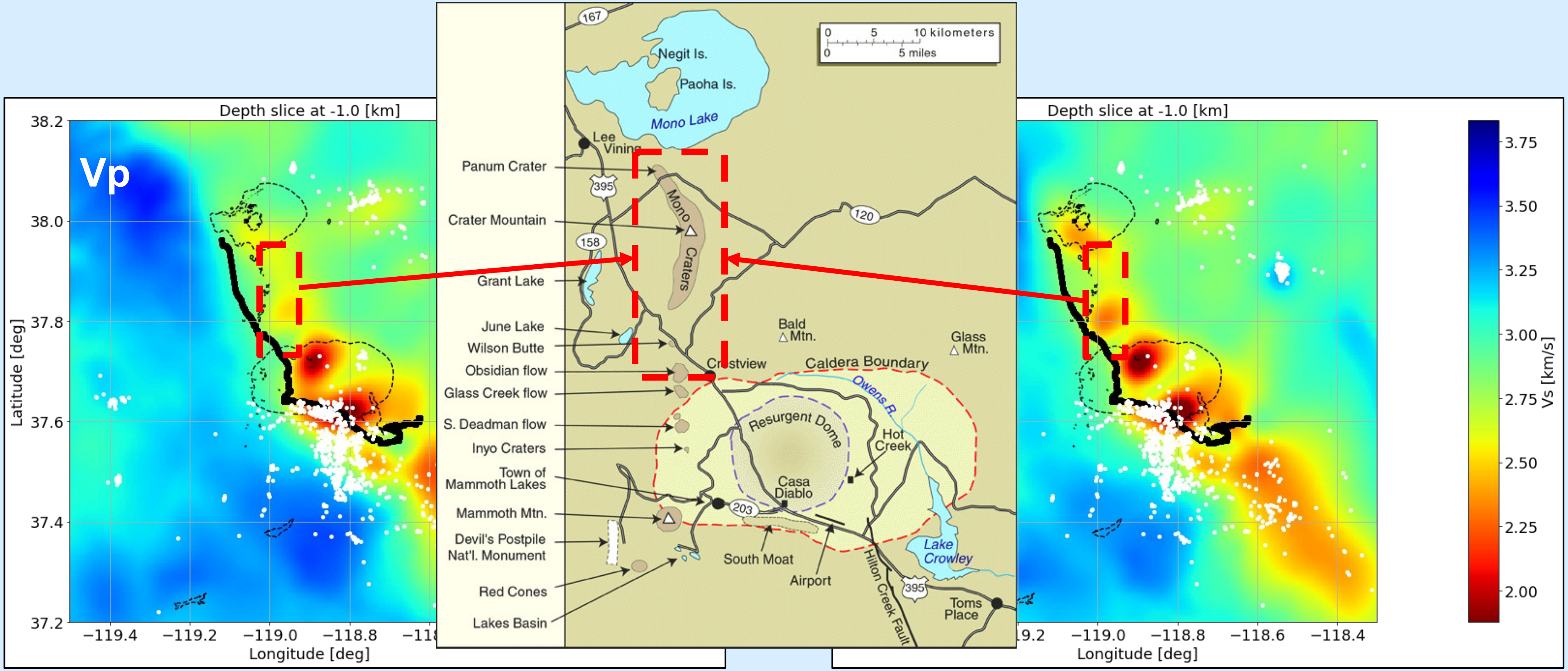
The Long Valley caldera: DAS tomography

Inverted model

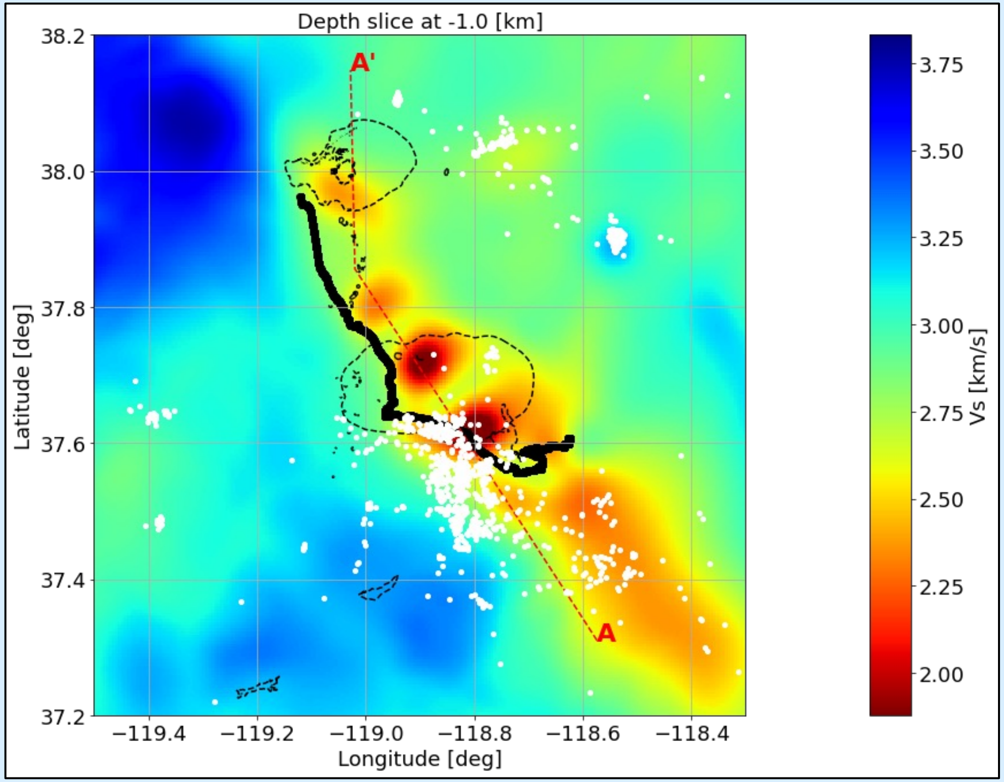




The Long Valley caldera: DAS tomography

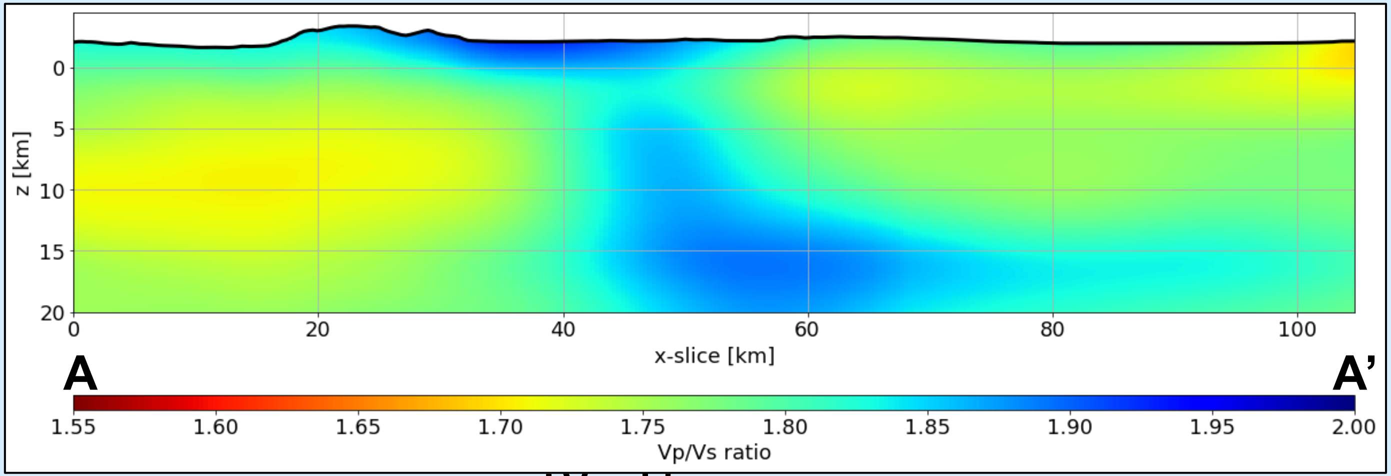


The Long Valley caldera: DAS tomography

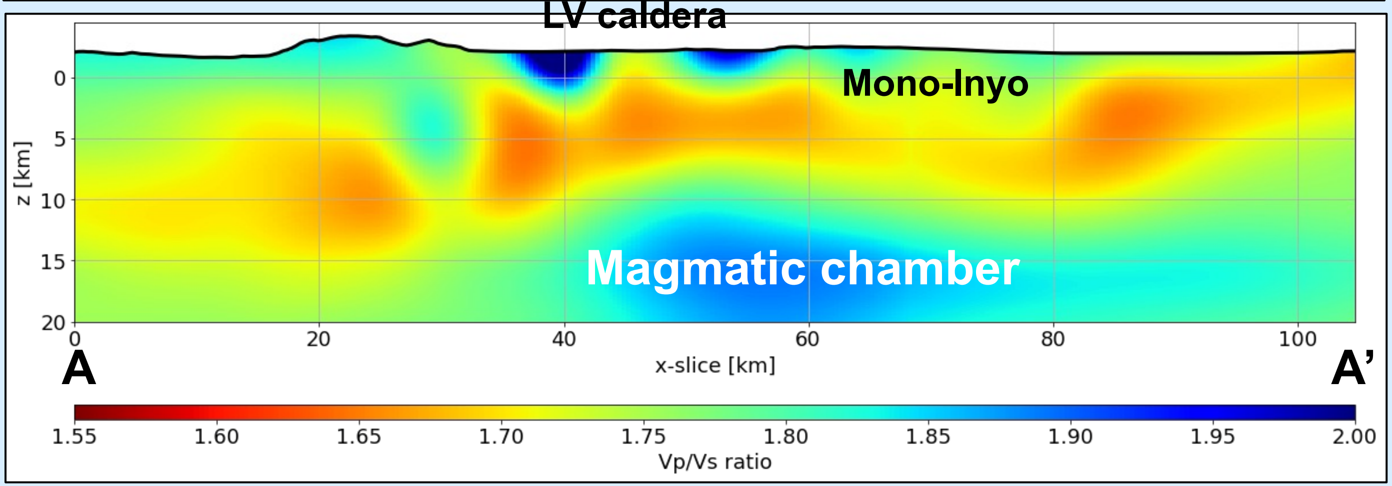


The Long Valley caldera: DAS tomography

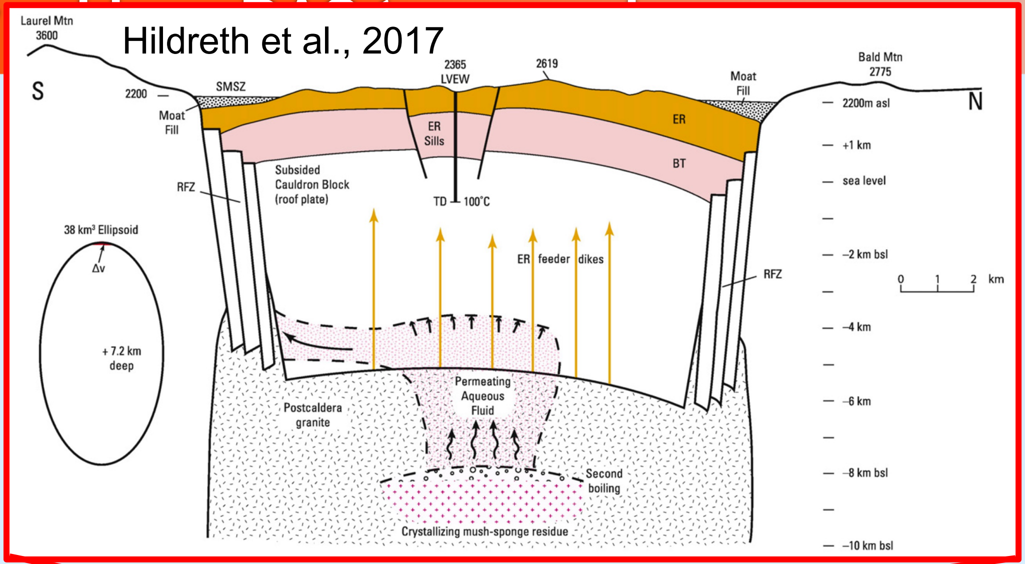
Initial Vp/Vs



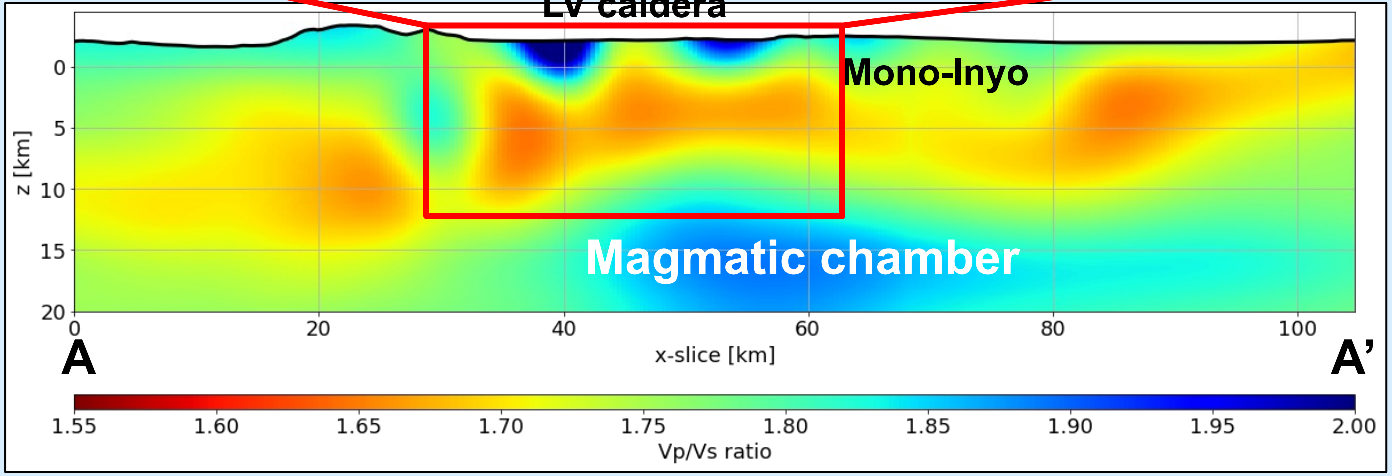
Inverted Vp/Vs



The Long Valley


Hildreth et al., 2017

Inverted Vp/Vs



Conclusions

- DAS provides ultra-dense spatial arrays recording seismic signal with unprecedented level of details. However, DAS data volumes represent a novel challenge for the seismology community
- We are taking a bottom-up approach in which we learn how to deal with this challenge by solving science problems
- Proper leveraging of modern architectures and computational tools are making DAS an incredibly resourceful tool
- Such projects are helping identify relevant DAS portions to design compression and selection algorithm for long-term storage

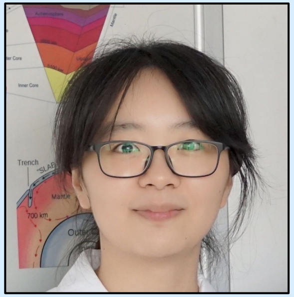
Acknowledgements

- **We would like to thank OptaSense for the constant support; in particular, we thank Martin Karrenbach, Victor Yartsev, and Vlad Bogdanov.**

- **We also thank the California Broadband Cooperative for providing access to the Digital 395 telecommunication fibers.**

- **We also thank the SCEC organizers for the invite to present our work.**

Acknowledgements

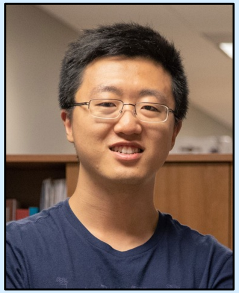


Yan Yang

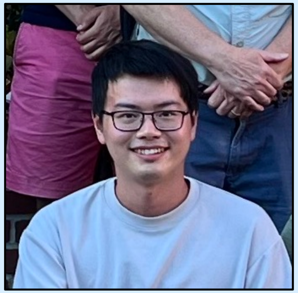
Jiaqi Fang

Ethan Williams

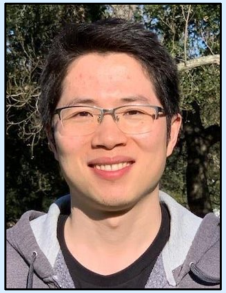
James Atterholt



Jiaxuan Li



Jiuxun Yin



Weiqiang Zhu

SCEC Southern California Earthquake Center
AN NSF+USGS CENTER

Thank you for your attention!

April 14th, 2022