DISTRIBUTED COMPUTING IN SENSOR NETWORKS

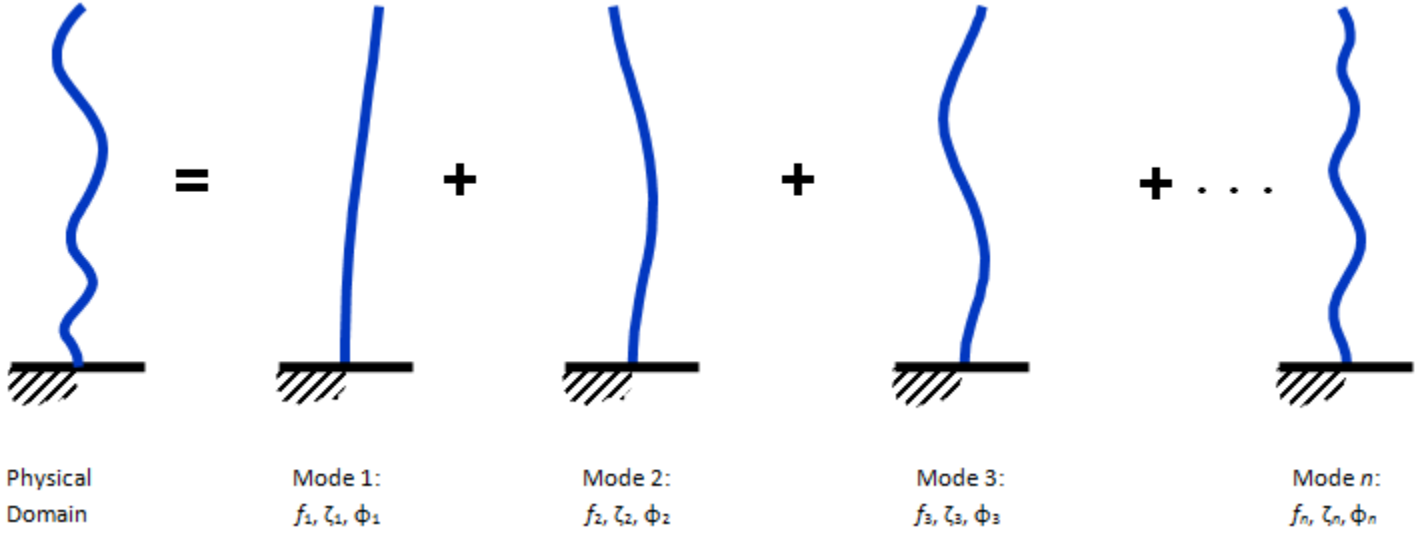
James Long [1] Anupa Murali[2] Zihao Wang [3]

[1]PhD Candidate
Civil Engineering
MIT

[2]Affiliation
Harvard University

[2]Affiliation
Harvard University

BACKGROUND



- Mode shapes describe the dynamic response of mechanical systems to their environment
- Material properties (mass, stiffness, damping), can be inferred from these responses
- Changes in modal responses may indicate defects

COMPUTATION

Given time series $x(t)$ from n sensor locations:

Compute $X_n(\omega)$ for each of the n locations

For a specific ω_i form $G_{n\times n}(\omega_i)$ where:

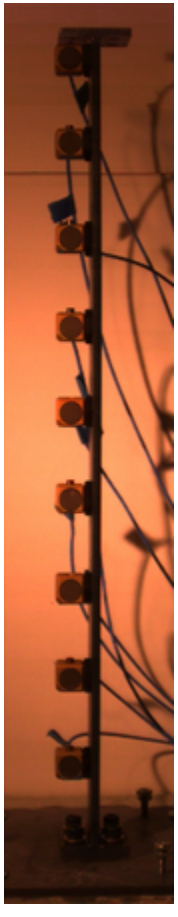
$$G_{jk}(\omega_i) = X_j(\omega_i) \times X_k(\omega_i)^*$$

Then decompose $G_{n\times n} = U_{n\times n}\Sigma_{n\times n}V^*_{n\times n}$

The first singular vector (column) of U is the mode shape estimate

MOTIVATION: TECHNOLOGY CAN OVERWHELM TRADITIONAL ALGORITHMS

Example 1: high-res cameras can produce huge data sets (e.g one time series per pixel)



MOTIVATION: TECHNOLOGY CAN OVERWHELM TRADITIONAL ALGORITHMS

Example 2: Networks of wirelessly connected sensor devices

- Devices have built in accelerometers and microcontrollers
- Very low cost enables large scale, dense deployment
- Communication out of the network drains battery, and we are forced to perform computation on slow (100MHz), memory constrained (128kB RAM) devices

APPROACH & SECRET WEAPON

- Recall that we use only the first singular vector as an estimate of the mode shape
- Because our matrix is normal, we can use a power method to estimate the largest eigenvector (PageRank!)

Algorithm:

$$i = 0; \text{init } v_0$$

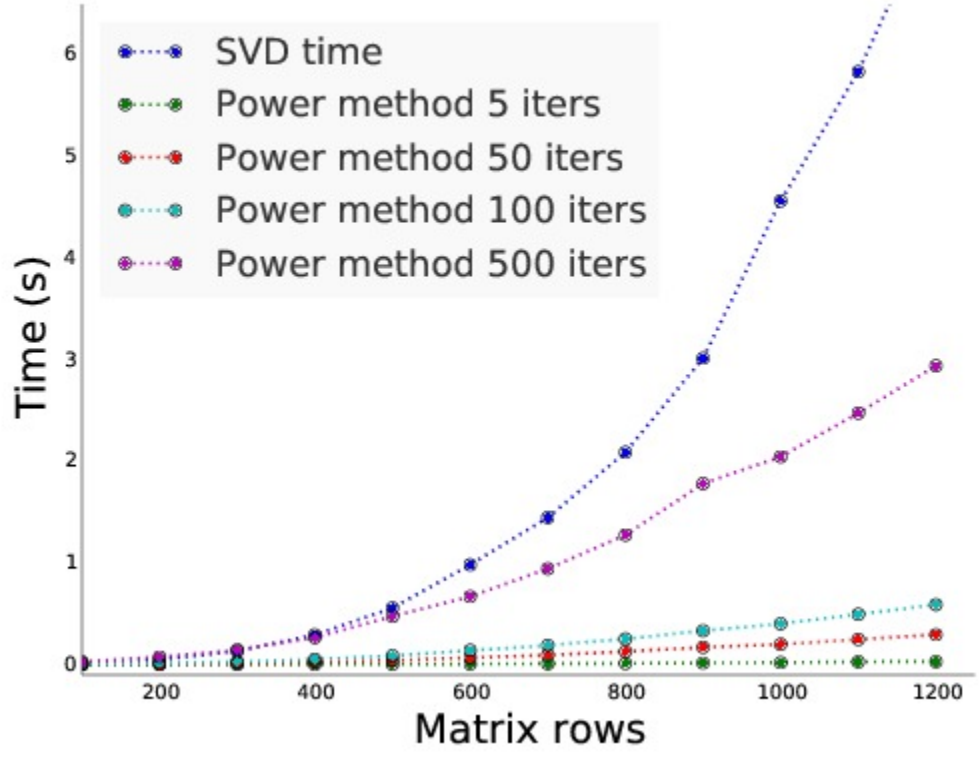
$$\text{while } (i < N) :$$

$$v_{i+1} = G v_i$$

$$v_{i+1} = \frac{v_{i+1}}{\|v_{i+1}\|}$$

$$i = i + 1$$

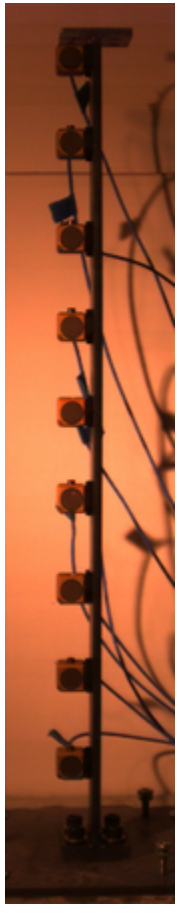
APPROACH & SECRET WEAPON

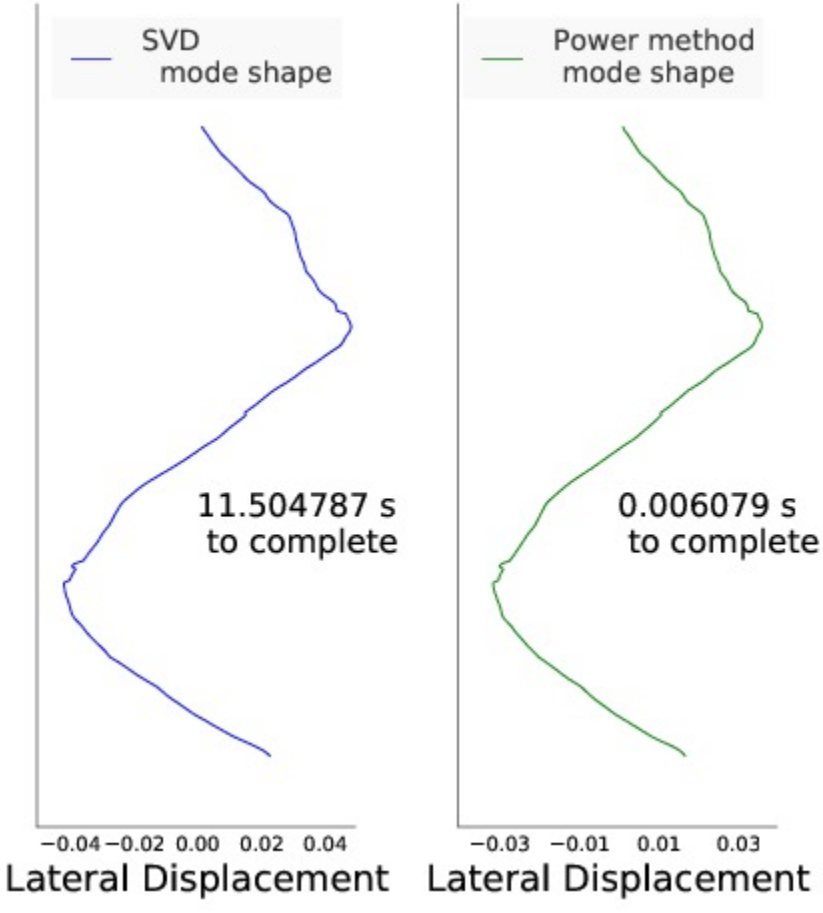


Rate of convergence: $\frac{\lambda_1}{\lambda_2}$

λ_1 is the largest eigenvalue, λ_2 the next largest

RESULTS: SERIAL COMPARISON



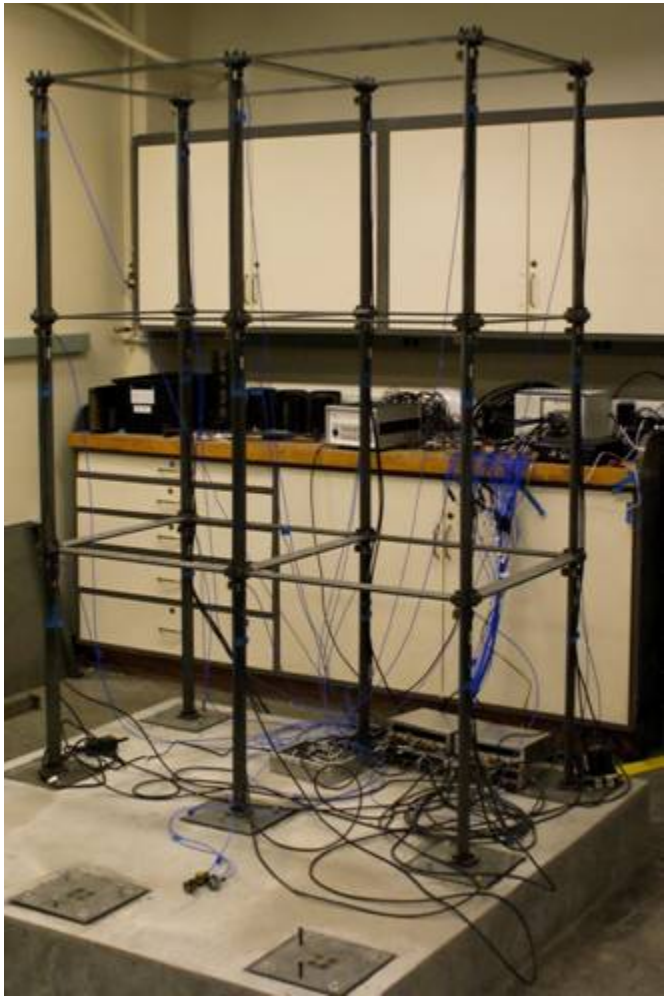


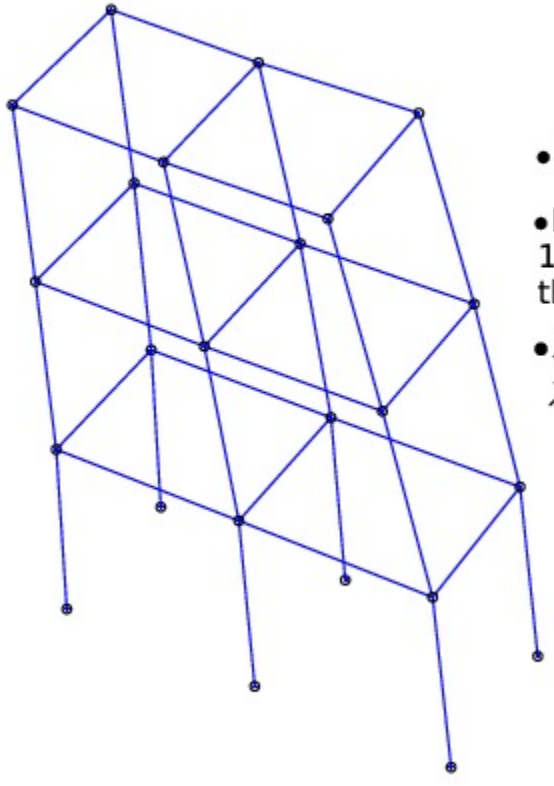
1217 individual time series of length 1000

1900x speed increase

$$\lambda_1 = 3.9 \times 10^5 \; \lambda_2 = 1.2 \times 10^{-9}$$

RESULTS: SERIAL COMPARISON





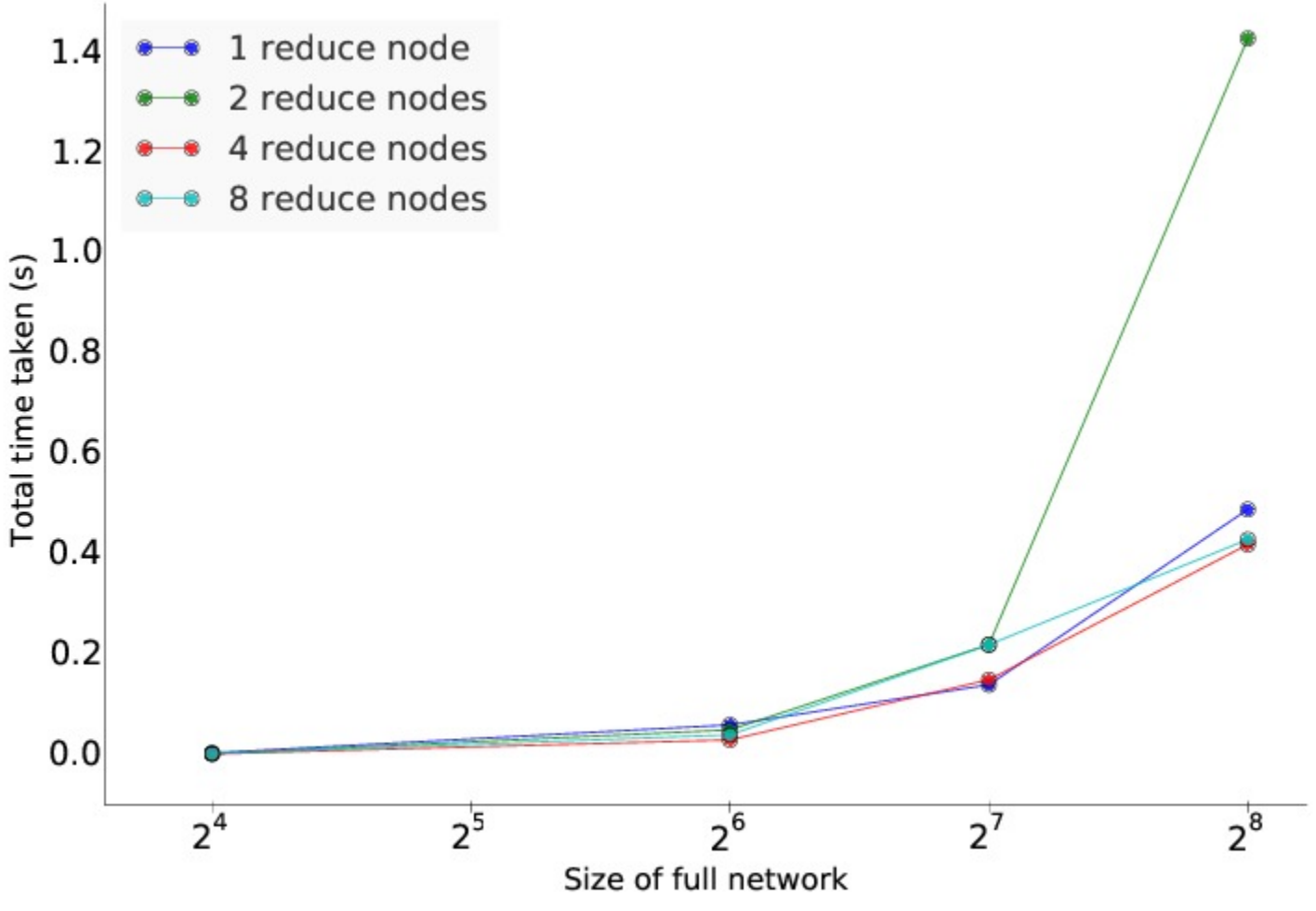
- 8 Hz mode
- Power method 10x quicker than SVD
- $\lambda_1 = 3.06 \times 10^{5}$
  $\lambda_2 = 5.85 \times 10^{-11}$

54 individual time series of length 180000

RESULTS: PARALLELISATION IN MPI

- MPI offers a solution to very large data sets, as well as a good simulation of a sensor network (using send/recv)
- We can exploit embarrasing parallelism of FFT calculation: One process for every time series
- Then collect FFT values at a subset of processes ('reduce'), otherwise communication will kill us
- Question: What is the right number of 'reduce' nodes

RESULTS: MPI ODYSSEY BENCHMARKING



Note: A realistic serial implementation would be orders of magnitude more costly

EVALUATION

- Using the power method to estimate mode shapes, instead of SVD saves orders of magnitudes in execution time on CPU (1900$\times$ on 1217$\times$ 1217 matrix)
- No meaningful decrease in accuracy
- Succesfully parallelised the algorithm using MPI4py: Simulation of a wireless network
- More work required to prove concept with wireless communication and microcontroller

MAIN LESSON

- Parallelism can help immensely - or hurt

- We need to decide prudently how much parallelism to use (MapReduce is a good example)

- Arithmetic intensity is important!

- Good algorithms are crucial