Differentiable Scientific Computing for ML

Ramani Duraiswami

Bowen Zhi, Armin Gerami, Meenakshi Krishnan, Leslie Li, Alisha Sharma





UNIVERSITY OF MARYLAND

IN THE NEWS | October 29, 2014

VisiSonics' RealSpace 3D Audio Software Licensed by Oculus for Virtual Reality

Personal Story

Developed/commercialized 3D Spatial audio rendering and personalization technology

Worked on wave propagation, vision, audition, FMM, parallel computing, statistics

Missed deep learning

Exited in 2023

Since Get into learning based computational physics

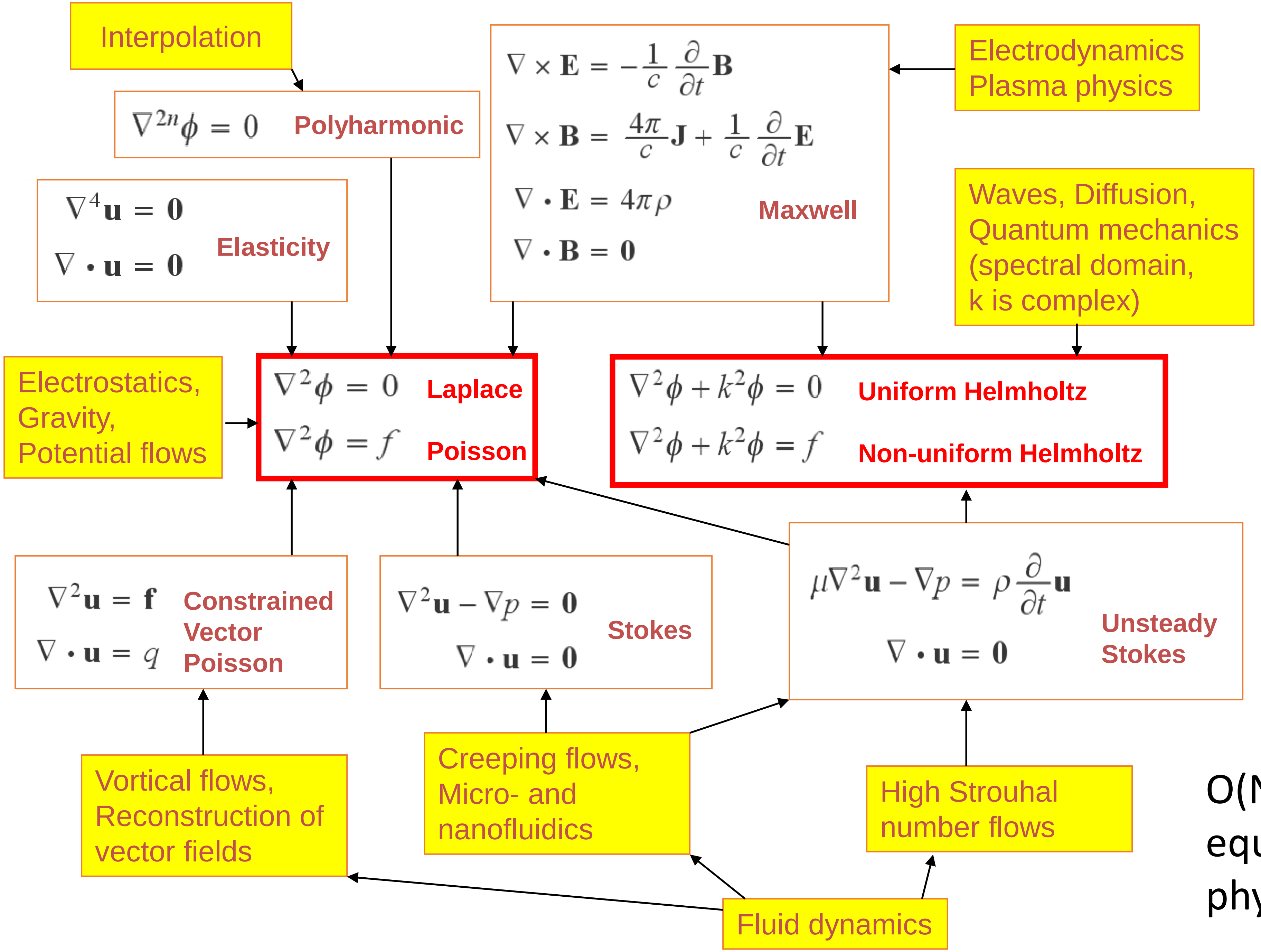
< BACK

VisiSonics

BLOG, EVENT, IN THE NEWS | May 10, 2023

CEVA Acquires Spatial Audio Business from VisiSonics to Expand its

Fast Multipole Methods



Application of FMM-BEM solvers to other domains

- Electrostatics
- Low-freq. / high-freq. Maxwell
- Fluid dynamics

O(N) Algorithms for classical equations of mathematical physics. Support AD

Two Themes for Current Research

Accelerating Deep Learning

- FMM and Parallel Computing
- O(N) Transformer -> on arXiv

Using ML and Differentiability to tackle problems in Physics and Engineering

- This talk

Aside – a recently submitted paper on arXiv

https://arxiv.org/abs/2402.07901

FASTFormer

Armin Gerami, Monte Hoover, Pranav Dulepet, Ramani Duraiswami

Perceptutal Interfaces and Reality Lab

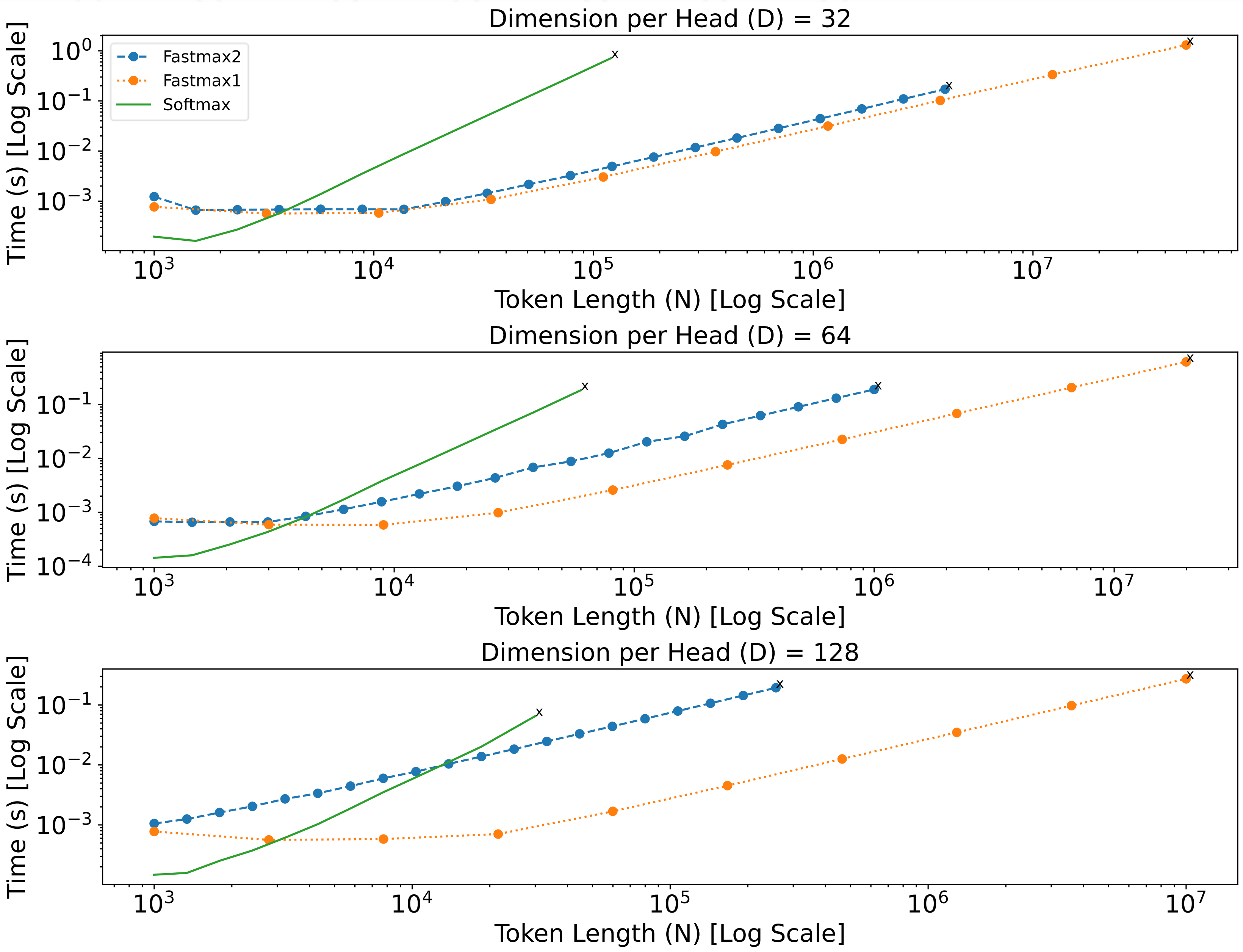
UNIVERSITY OF MARYLAND 1856

PIRL

Complexity

- Calculating the expansion terms require $O(ND^{p+1})$ time and $O(D^{p+1})$ memory, and calculating the output score $O(ND^{p+1})$ time and $O(ND^{p})$ memory, totaling to $O(ND^{p+1})$ time an $O(ND^{p}+D^{p+1})$ memory (p is the Taylo expansion length).

- The plots show time taken for a single forward pass on A6000. Notice how Vanilla Attentions (softmax) scales quadratically with N, whereas we scal linearly.



Expressivity

- To measure the expressivity, we use the Long Range Arena (LRA) benchmark. As we can see, our speedup does not introduce a compromise in terms of expressivity.

Model	ListOps	Text	Retrieval	Image	Pathfinder	Avg
Vanilla Trans.	38.37	61.95	80.69	40.57	65.26	57.37
Informer	36.95	63.60	75.25	37.55	50.33	52.74
Reformer	37.00	64.75	78.50	43.72	66.40	58.07
Linear Trans.	16.13	65.90	53.09	42.34	75.30	50.55
Performer	37.80	64.39	79.05	39.78	67.41	57.69
Fastmax2 (ours)	37.40	64.30	78.11	43.18	66.55	57.90
Fastmax1 (ours)	37.20	63.25	78.21	42.76	66.67	57.62

- Our speedup is evident in LRA

Model	ListOps ($N = 2000$)	Text ($N = 4000$)	Retrieval ($N = 4000$)	Image ($N = 1000$)	Pathfinder ($N = 1000$)	Avg
Vanilla Trans.	6.4	1.8	1.7	3.0	6.1	3.8
Fastmax2 (ours)	11.6	6.1	6.9	3.0	6.8	6.9
Fastmax1 (ours)	47.4	26.7	24.7	12.8	24.4	27.2

Differentiable Scientific Computing for ML

Ramani Duraiswami

Bowen Zhi, Armin Gerami, Meenakshi Krishnan,



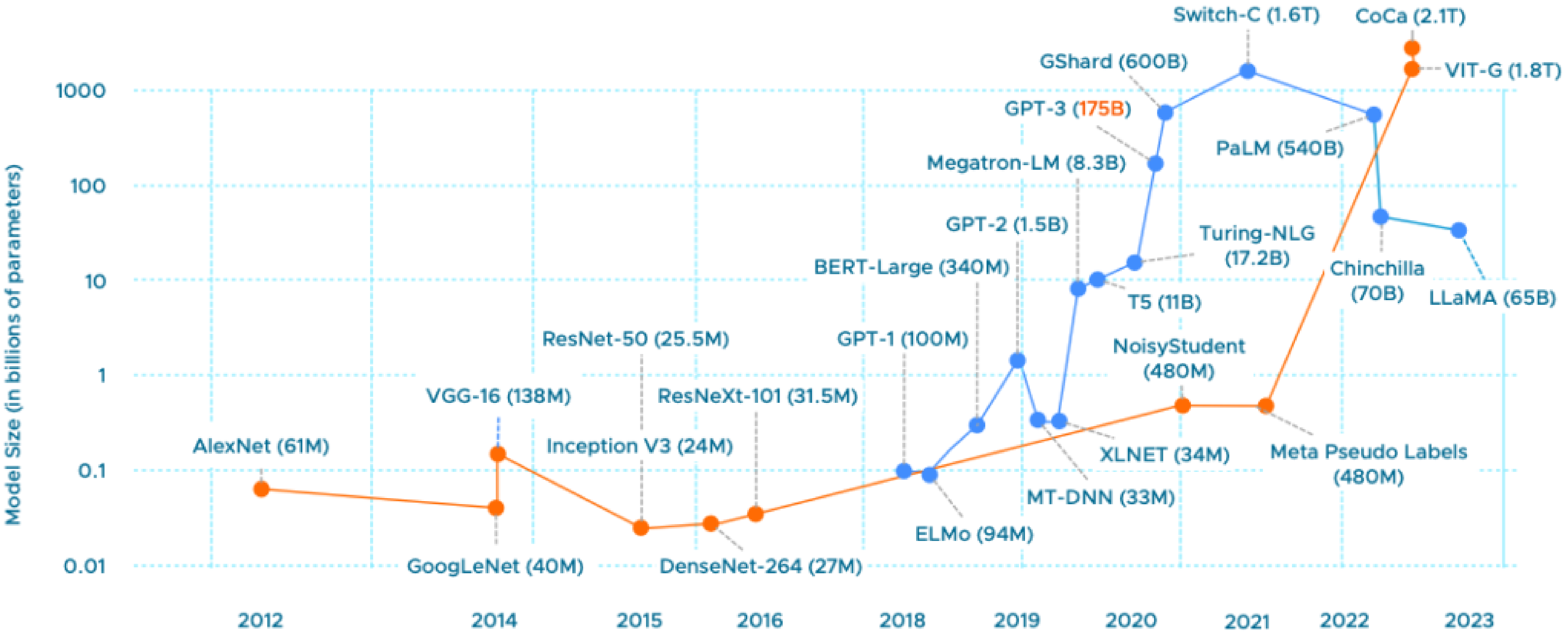
Leslie Li, Alisha Sharma


UNIVERSITY OF MARYLAND

Deep Learning

- Incredible progress in data rich domains
  - Images, Speech, Vision, Language
  - Multiple such modalities
- Can capture
  - Functions: Predict expected values from data - Input-output relations
  - Distributions: Generate plausible outputs
- Goal: Figure out how to use deep learning in settings where there are
  - strong “forward” models
  - data maybe sparse
- Not just use the forward solver to build datasets for learning

Deep network size



Pillars of Scientific Discovery

Explainable

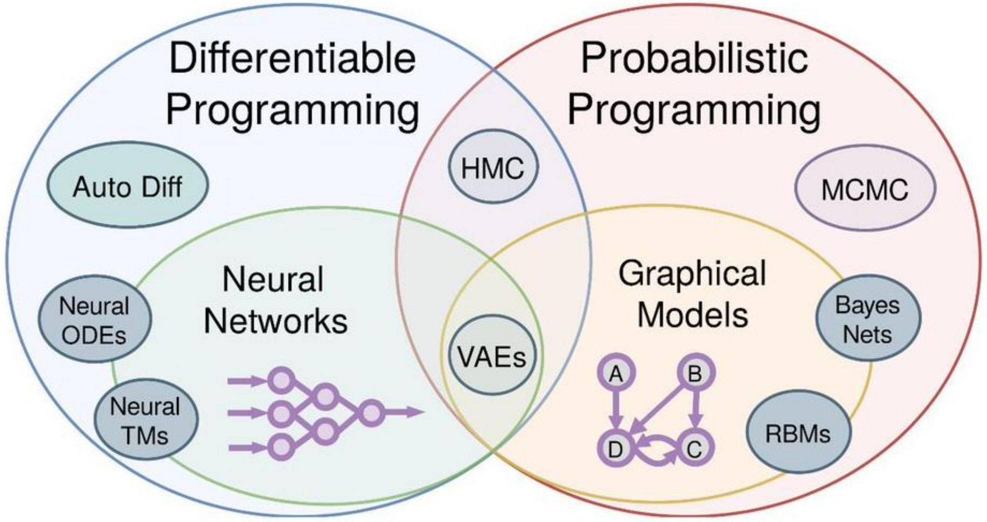
Experiment /Hypothesis	Modeling	Computational Simulation	Data
• explain observations • "Scientific Method"	• Mathematical Physics to generalize • Equations and initial/boundary conditions • generalizes	• Scientific Computing and Numerical Analysis • Further generalize and explore • Inverse Problems	• Learn from Data • Deep Learning • Generalize • Not yet very explainable

Back Propagation, the Computational Graph

- A less appreciated component of the DL revolution
  - Automatic differentiation
- Simple compositional network models used in DL can be automatically differentiated
- Parameters determined via optimization
  - Stochastic Gradient Descent and ADAM flavors
- Even a high school student can build a cost function computation code and then have their DL model trained
  - Tensorflow, Pytorch
- Gives deep learning folk the feeling any problem can be fit in
- How to fit our forward models into this framework?

Relevant Research Streams

- Automatic Differentiation - classical field, but very active in compilers and ML; Particularly appreciated C. Elliott
- Differentiable Programming – several frameworks to build gradients automatically
- Differentiable Physics – many researchers, but found work of Nils Thuerey very accessible
- Neural Radiance Fields – very vibrant field; has taken over computer graphics; starting from Mildenhall et al 2020.

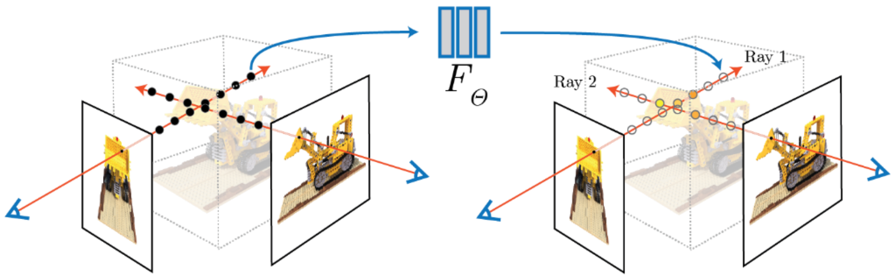


M.Sc. Thesis, Considine, 2020

Google JAX - 2018

PyTorch - 2019

Julia SciML - 2017

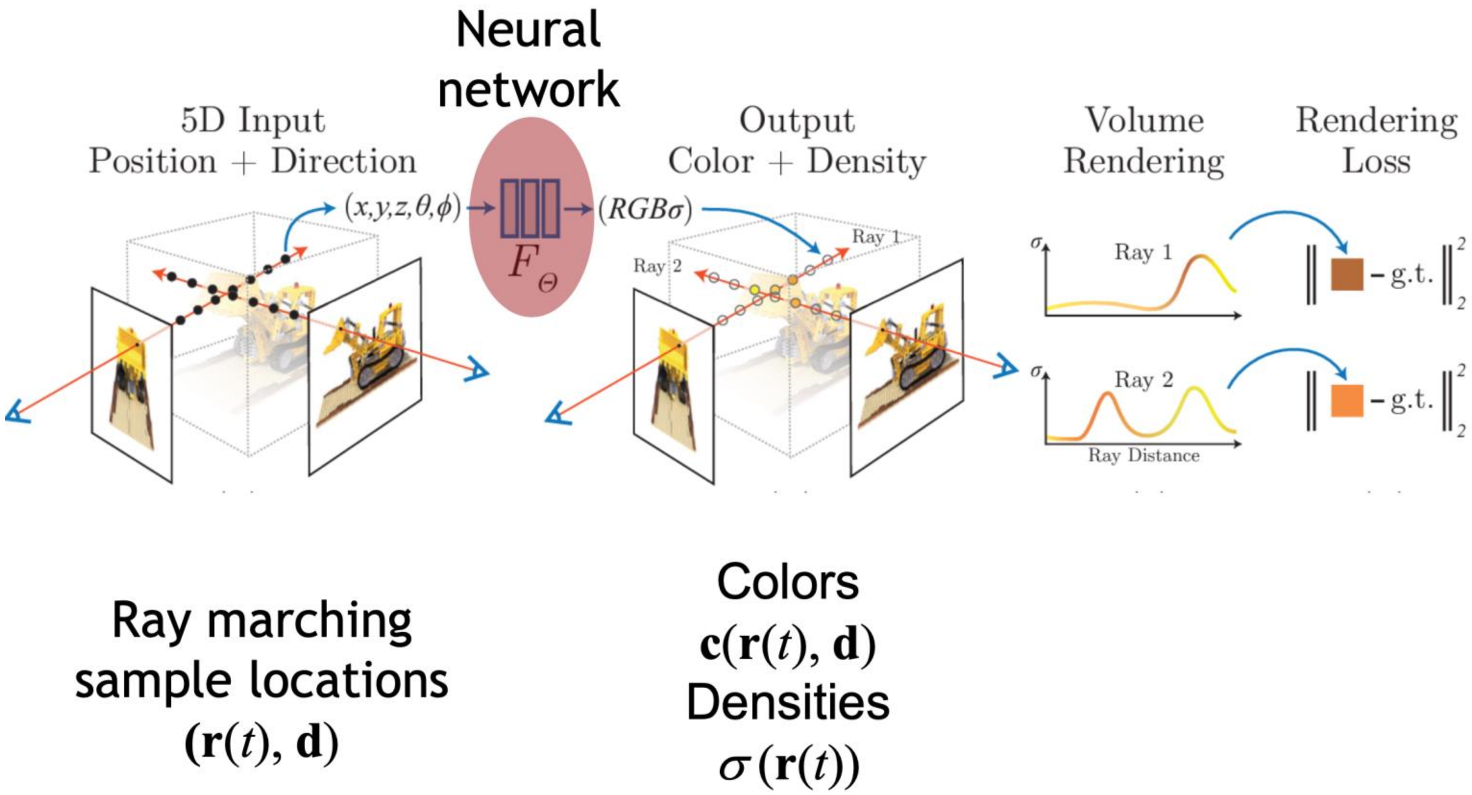


Neural Radiance Fields

PBDL

NeRF

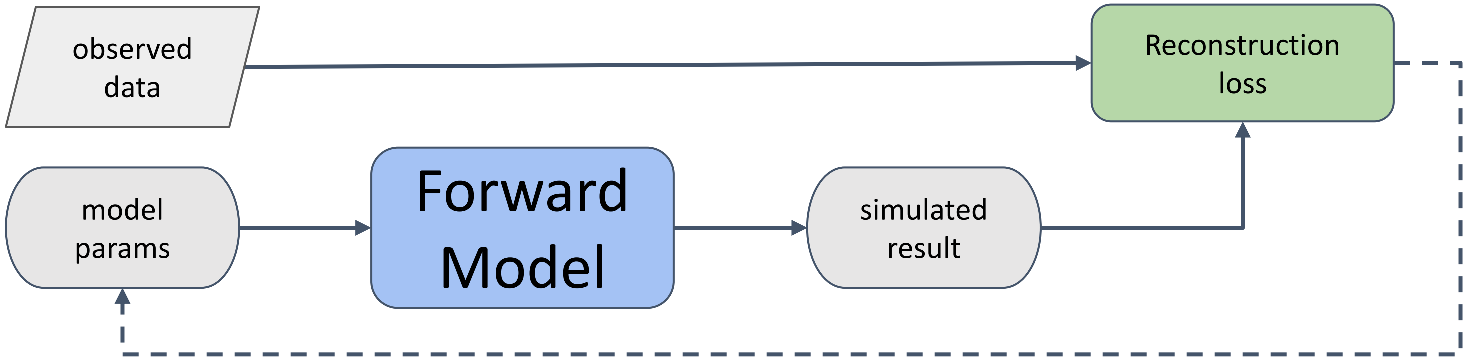
- **Given:** a rendering equation and snapshots of the scene from different angles
- **Objective:** 3D scene representation
  (i.e. learned functional representation of volumetric density, point color)
- Applications: 3D geometry extraction/Novel view synthesis
- Backpropagation – differentiation through forward solver



arxiv.org/abs/2104.06405

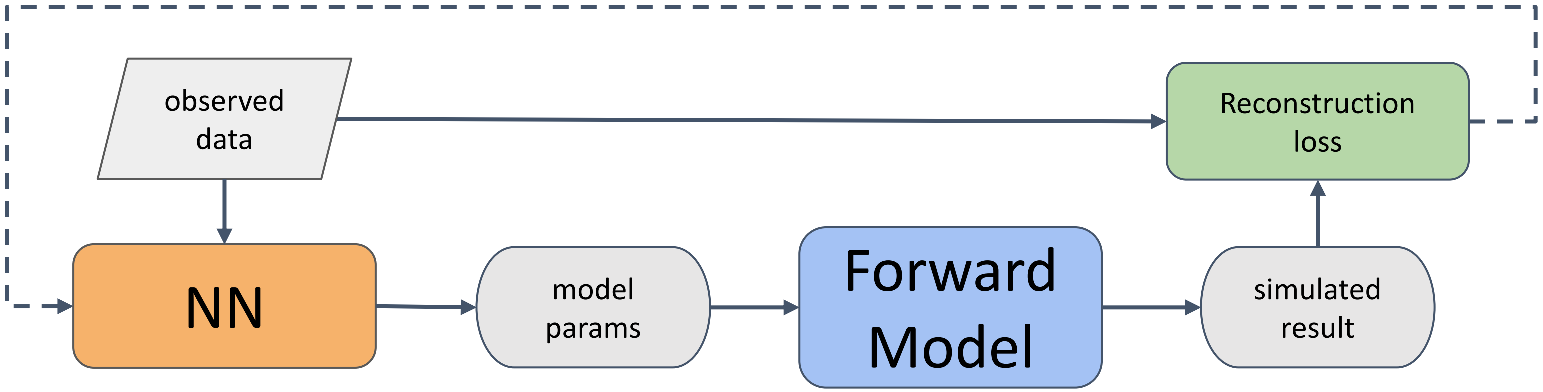
Differentiable Physics – beyond creating data

- Gradient-based inverse problem optimization
  - Possible with a differentiable physics-based forward model of the problem



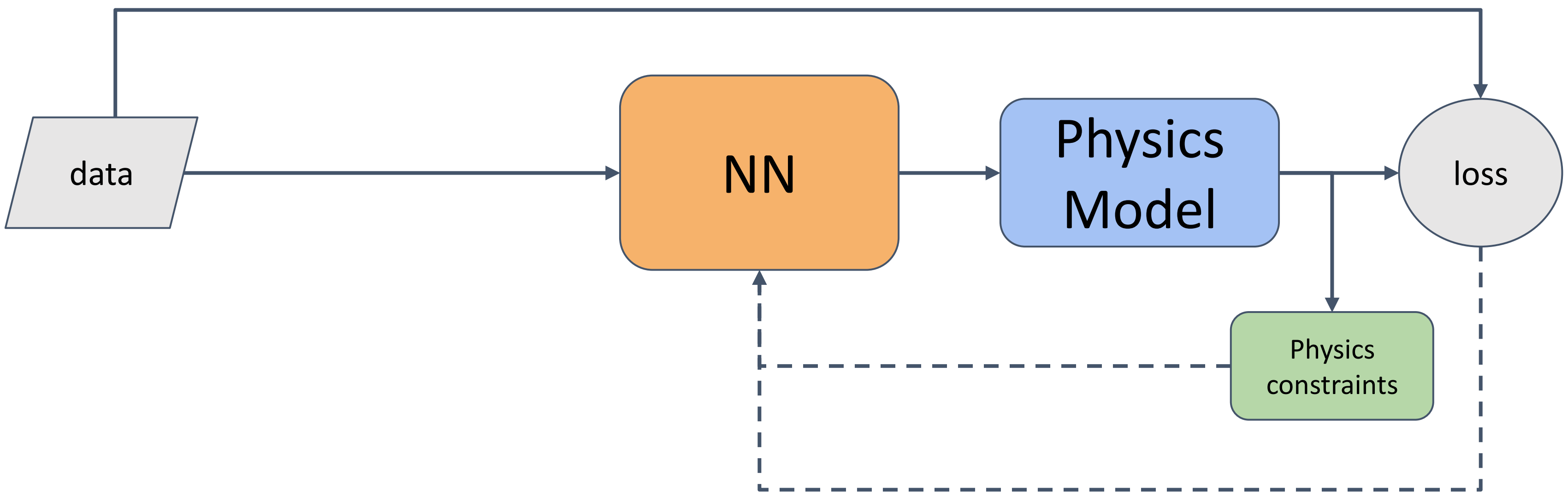
Differentiable Physics – beyond creating data

- Gradient-based inverse problem optimization
  - Can use a trained NN to better condition the problem with learned priors from a dataset



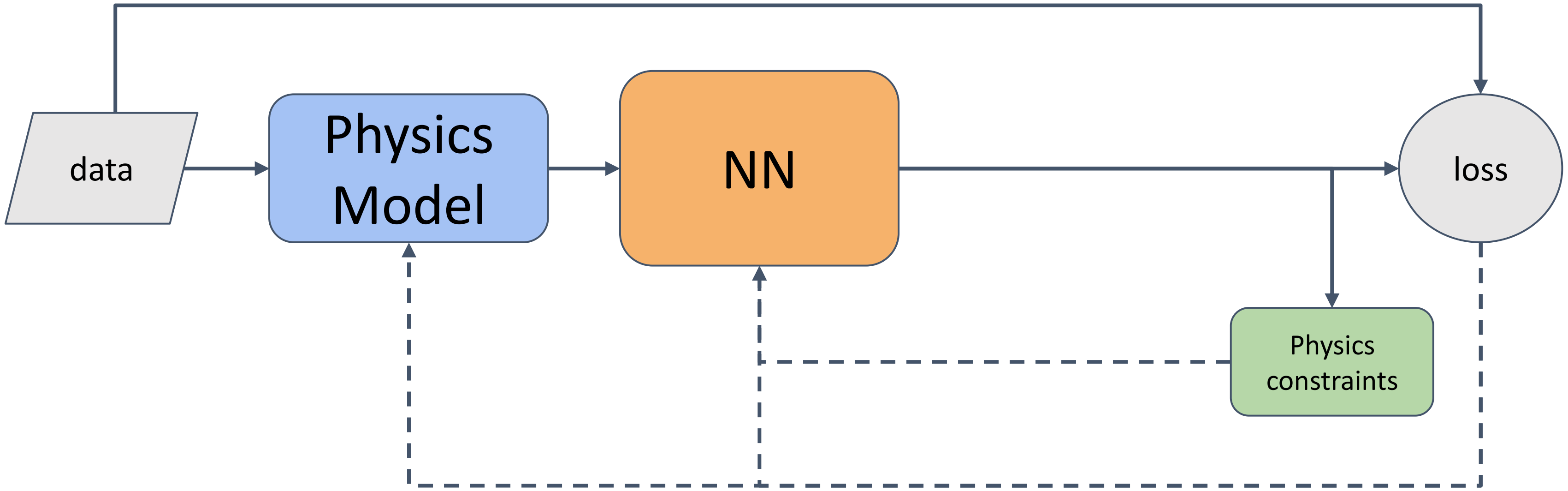
Differentiable Physics – beyond creating data

- Gradient-based inverse problem optimization
  - Can use a NN to suggest which Physics model should be used!
  - Perhaps trained via LLMs
- Many possible variations



Differentiable Physics

- Gradient-based inverse problem optimization
  - Use the data and process it with a physics-based forward model
  - Have a trained NN which learns the predictions of the physics model
- Many possible variations



Initial Projects

- Differentiable
  - Electrical Impedance Tomography
  - Room Acoustics Models
  - DSP Filters
  - Models of Human Hearing

Example: Electrical Impedance Tomography

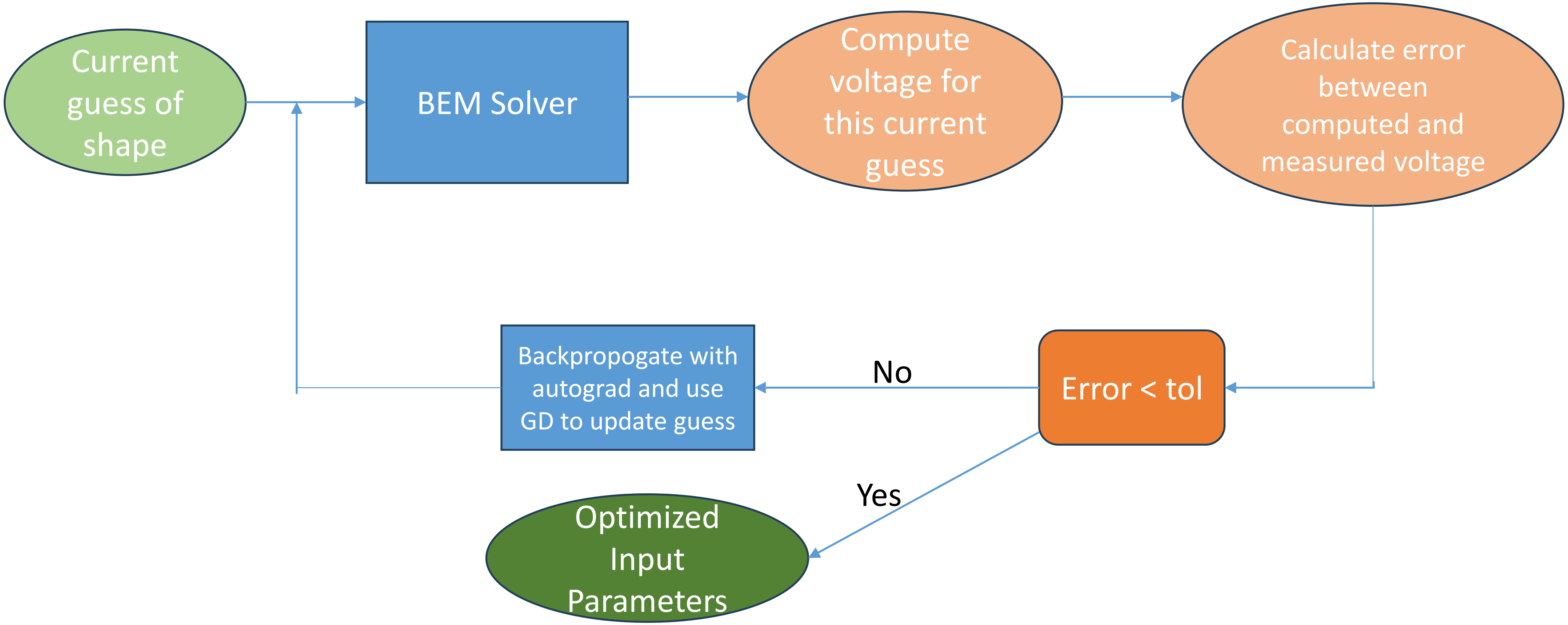
- Problem: Reconstruct the distribution of impedances inside an object ("image") by applying current at some electrodes and taking measurements of the voltage at other electrodes.
- Examples: Imaging gas bubbles in a liquid, cracks in a conducting material etc.
- The electric potential Φ satisfies the Laplace equation where **n** is the outward normal to the boundary and σ is the electrical conductivity:

$$\nabla.(\sigma\nabla\phi) = 0 \text{ in } \Omega \subset \mathbb{R}^n$$

$$\text{subject to } \begin{cases} \frac{\partial\phi}{\partial n} \text{ and } \phi \text{ known at electrodes,} \\ \frac{\partial\phi}{\partial n} = 0 \text{ everywhere else.} \end{cases}$$

- In applications with regions of vanishing conductivity., the task becomes to identify the shapes of these regions.
- Boundary Element Method – A good choice of forward solver in these problems.

Duraiswami et al. "Boundary element techniques for efficient 2-D and 3-D electrical impedance tomography." *Chemical engineering science* 52.13 (1997): 2185-2196.


Current guess of shape
BEM Solver
Compute voltage for this current guess
Calculate error between computed and measured voltage
Error < tol
No
Backpropogate with autograd and use GD to update guess
Yes
Optimized Input Parameters

Many issues with practical implementation

- Started with trying to do our FMM BEM Solvers
- AD does not work well for complex problems
  - Memory size grows very quickly for Scientific Computing
- Was finally able to solve this
- Many operations are not obviously differentiable
  - Meshing
  - Boolean decisions
- Linear Algebra beyond BLAS not yet differentiable
  - Matrix decompositions
- Part of our longer-term research
- Indeed, Diff Physics folks mostly solve Burger's equation!
- And NERF folks use a stochastic model!

How do we bring numerical analysis back into solving problems

Consider simplest linear least squares regression of

$$Ax = b \quad \text{with the cost function} \quad ||b - Ax||_2$$

Automatic differentiation frameworks would be unable to come up with the normal equations

$$A^t A x = A^t b$$

Any self-respecting numerical analyst use rank-revealing QR or regularized SVD.

JAX will not be able to learn these from AD

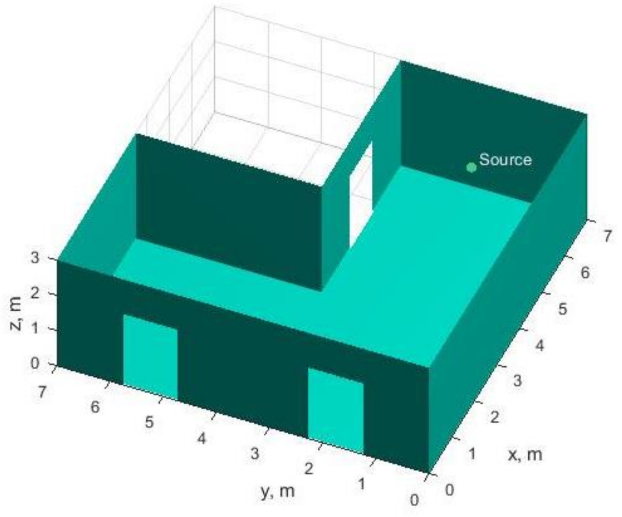
A Differentiable Image Source Model for Room Acoustics Optimization

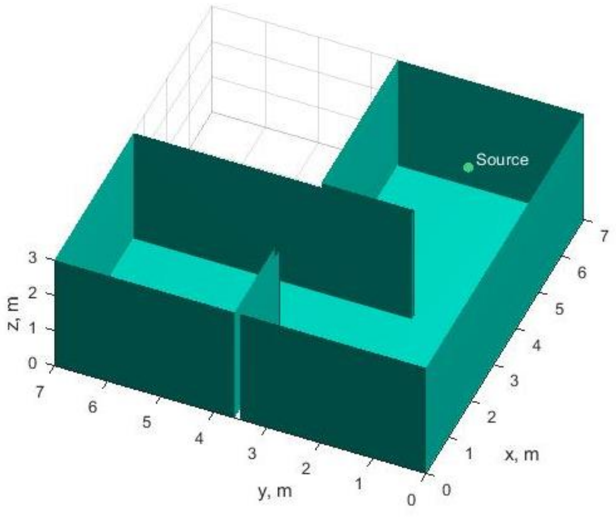
Bowen Zhi, Alisha Sharma, Dmitry Zotkin, Ramani Duraiswami

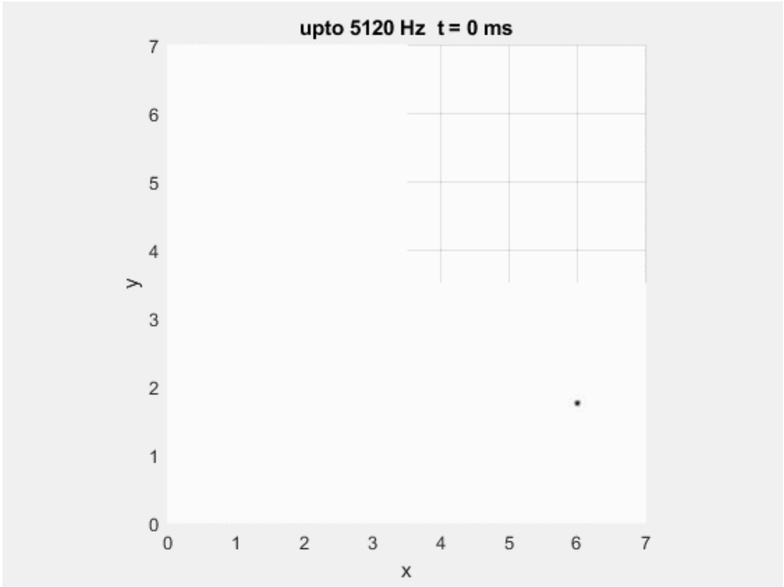
Perceptual Interfaces and Reality Lab,
Computer Science & UMIACS,
University of Maryland, College Park

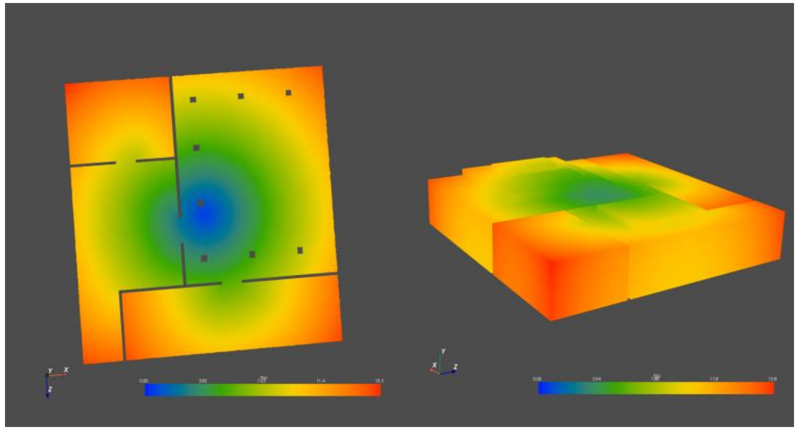
Work supported by ONR

Accurate room acoustics simulation with Helmholtz and Eikonal solvers







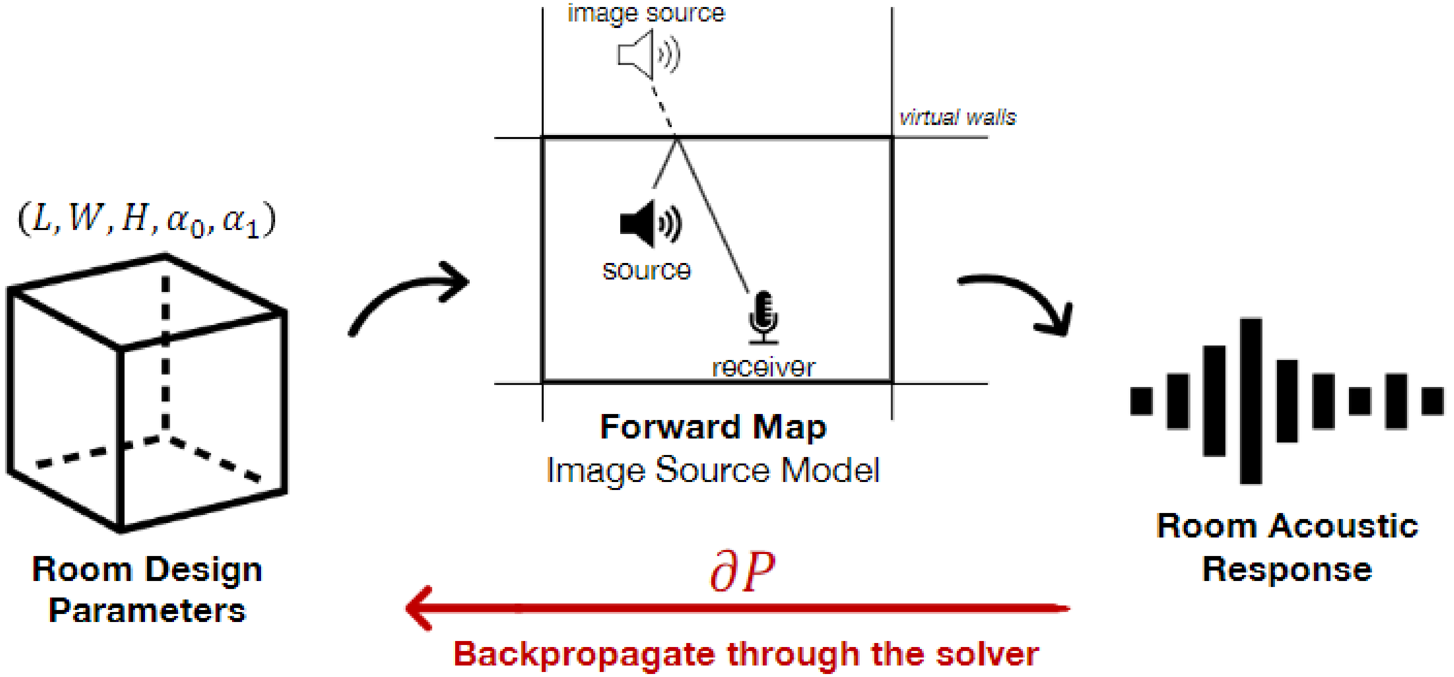


Potter et al., (2022)

- Helmholtz solvers:
  - Physically accurate full wave simulation
  - We have efficient algorithms and solvers
    - Including efficient GPU versions
- Eikonal solvers: “principled raytracing”
  - Unlike raytracing for graphics, acoustic raytracing lacks physical justification and cannot compute some important real physical effects, e.g. diffraction and scattering. Eikonal solvers address these points.

Differentiable Room Acoustics

- Differentiable image-source method
  - Implemented in PyTorch[5]
  - Input: room dimensions, wall materials, locations of source and receiver
  - Output: room impulse response $h(t)$
  - For cuboid rooms, can obtain gradient w.r.t. all inputs
    - Including geometry



[5] *Paszke et al., Neurips, 2019*

Differentiable Room Acoustics

- Differentiable image-source method
  - Implemented in PyTorch
  - Input: room dimensions, wall materials, locations of source and receiver
  - Output: room impulse resp $h(t)$e
  - For cuboid rooms, can obtain gradient w.r.t. all inputs
    - Including geometry
  - Simple formulation
    - Sum of delayed and attenuated impulses

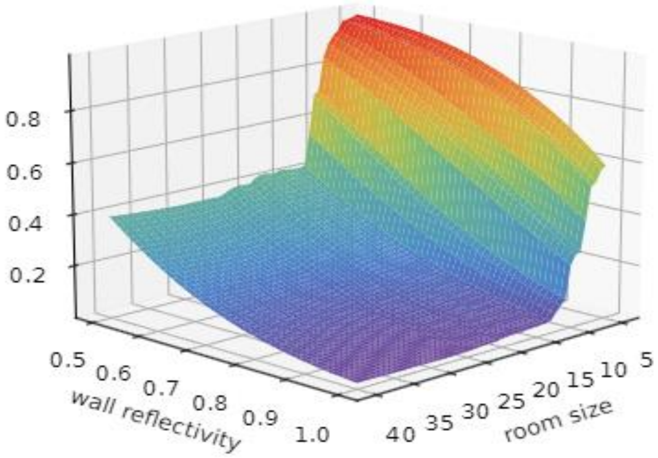
$$h(t) = \sum_k \left(\underbrace{\delta\left(t - d_k/c\right)}_{\text{Impulse function}} \underbrace{e^{-d_k/\xi}}_{\text{Air absorption}} \underbrace{\prod_{x \in W_k} \left(-\rho(x)\right)}_{\text{Wall material absorption}} \right), \quad d_k = \|s_k - r\|$$

Room Acoustics Metrics

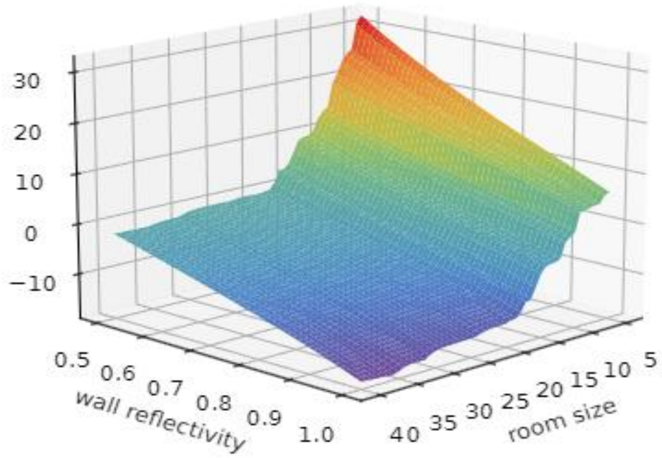
- Implemented several metrics commonly used in architectural acoustics[6]
  - Computed directly from the impulse response
    - Reverberation time T
    - Definition D
    - Clarity Index C_{80}
    - Center Time
    - Speech Transmission Index (STI)
- Question: can we use such metrics in gradient-based optimization problems?
  - Examples:
    - find room dimensions and wall materials to achieve target metric values
    - find optimal placements of speakers / microphones
  - Ideally: metrics are smooth, monotonic / convex w.r.t. scene parameters
  - Evaluate these metrics and their gradients using our acoustics simulator over a domain of varied scene parameters

[6] *Kuttruff, "Room Acoustics", 2016*

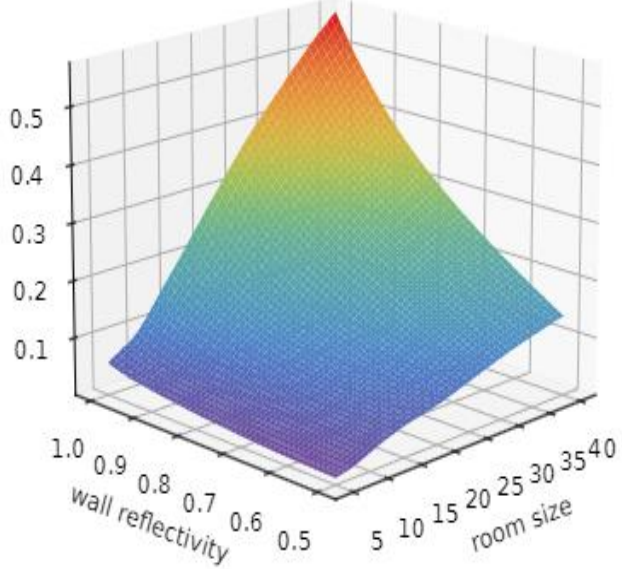
- Question: can we use such metrics in gradient-based optimization?



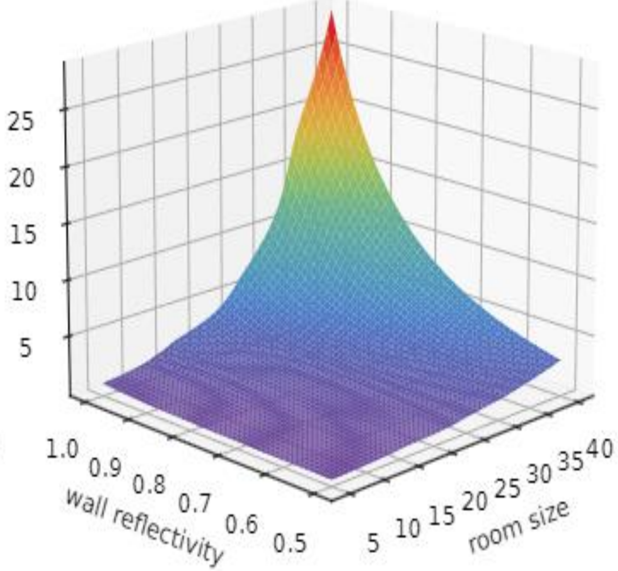
(a): Definition, D



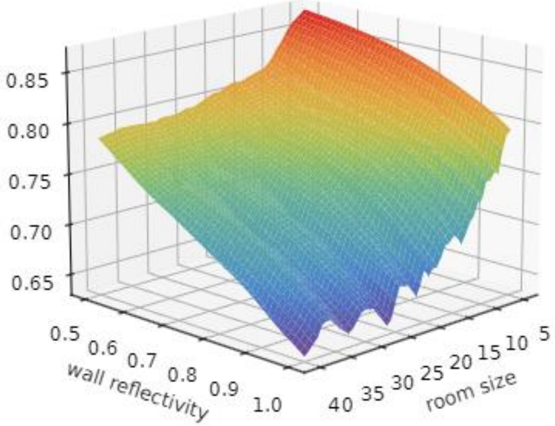
(b): Clarity Index, C_{80}



(c): Center Time, t_s



(d): Reverb Time, T



(e): Speech Transmission Index, STI

Optimization Experiments

- Example optimization problems
  - Find room dimensions s.t. when some number of walls have their materials changed, the difference in reverb time is maximized
  - Find the receiver location in the room that minimizes/maximizes acoustics metrics
    - E.g., placing a voice assistant optimally

N_D	1	2	3	4	5
$\|T_0 - T_1\|$	0.81	1.87	6.05	9.76	18.23

Table 1: Maximal difference in reverberation times (in s) achieved for each wall configuration. N_D denotes the number of walls whose properties can be dynamically adjusted.

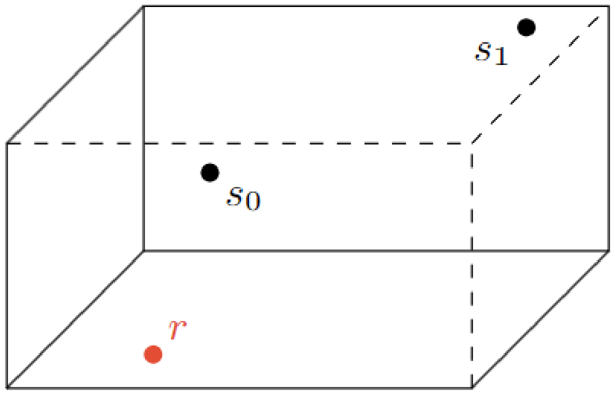


Figure 6: Visualization of the receiver location r obtained from minimizing the center times to s_0 and s_1.

Optimization Experiments

- Validation with real RIR data
  - MeshRIR dataset[7]
    - RIRs measured in a room at many source-receiver location pairs
    - Ground truth room dimensions, source-receiver locations
      - No precise information about room materials
  - Test problem: can we obtain reasonable estimates for room geometry based on the measured impulse responses?
    - Solve by comparing RIR metrics between observation and model output
      - Evaluate difference between metrics at various room parameter combinations
      - Sanity check: see if a (local) minima exists at the true room geometry for some feasible material values

[7] *Koyama et. al, WASPAA, 2021*

DIFFERENTIABLE FIR-TO-IIR FILTER ESTIMATION WITH APPLICATIONS

Armin Gerami, Bowen Zhi, Dmitry N. Zotkin, Ramani Duraiswami

{agerami, bzhi, wdz, ramanid}@umd.edu

Motivation

- Compared to FIRs, IIR filters are less computationally intensive, require less memory, and have lower latency.
- Our primary motivation was to make HRTF based scene rendering efficient for edge devices.
- Can also be used for efficient storage of HRTFs.
- Several other applications in DSP
  - Underwater Modems

Solution via Automatic Differentiation

- Given FIR filter with impulse response h, find IIR estimation $\hat{h}$, so that:

$$\hat{h} = \min_{k_i, a, i=[0,2N-1]} \mathrm{Cost}(h[i], \hat{h[i]}) \qquad \hat{h}[n] = \sum_{i=0}^{2N-1} k_i \alpha^{n-i} u[n-i]$$

- Novel Model Reformation – results in convex cost - can be used with AD

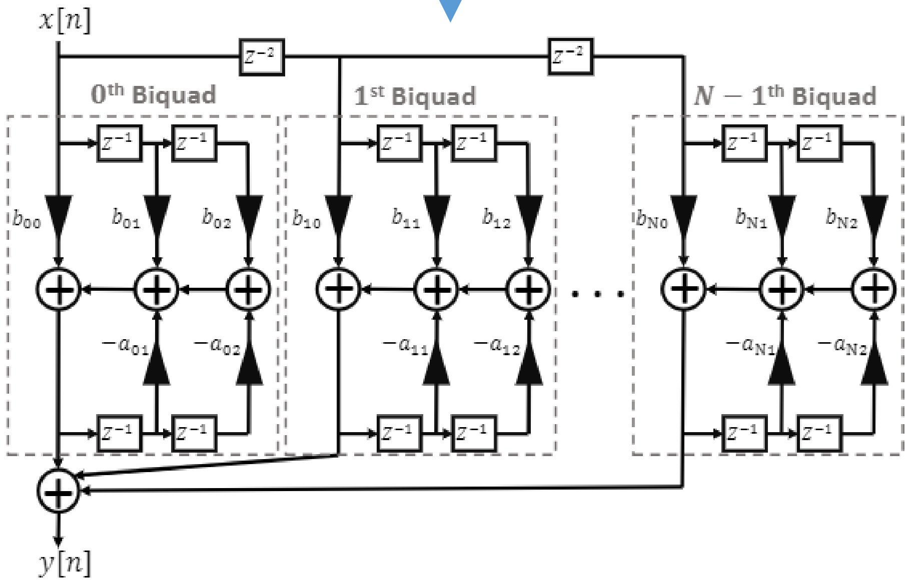
$$\sum_{i=0}^{2N-1} k_i \alpha_i^{n-i} u[n-i] \xrightarrow{\text{in z-domain}} \sum_{i=0}^{2N-1} \frac{k_i z^{-i}}{1-\alpha_i z^{-1}} \xrightarrow{\text{reshape}} \sum_{i=0}^{N-1} z^{-2i} \frac{b_{i0} + b_{i1} z^{-1} + b_{i2} z^{-2}}{1 + a_{i1} z^{-1} + a_{i2} z^{-2}}$$

(our model in time-domain)

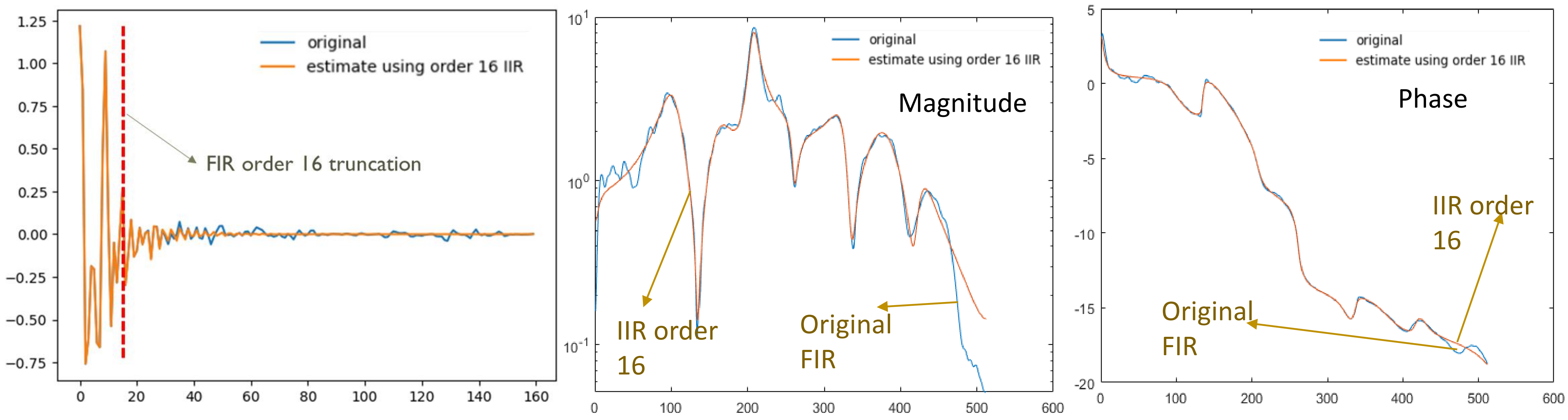
where

$$b_{i0} = k_{2i}, \quad b_{i1} = k_{2i+1} - k_{2i}\alpha_{2i+1} \quad b_{i2} = -k_{2i+1}\alpha_{2i},$$

$$a_{i1} = -\alpha_{2i} - \alpha_{2i+1}, \quad a_{i2} = \alpha_{2i}\alpha_{2i+1}$$



Application order 16 approximation of a 160 tap HRIR filter



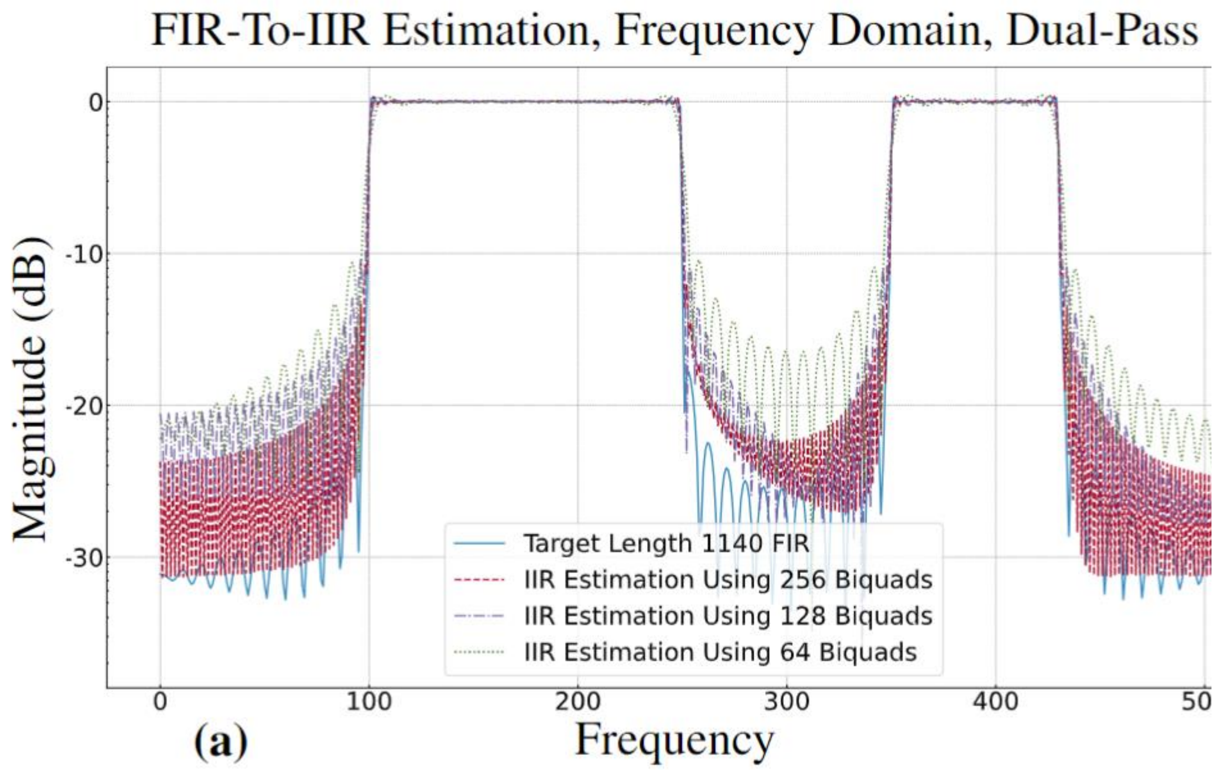
✓In time domain, the FIR onset matched almost exactly, while providing a good approximate of the tail
✓In frequency domain, desired spectral characteristics such as notches and peaks are preserved
✓Very accurate approximation of the phase, making our solution suitable for estimating digital filters as well.
✓**3x speedup; 5x memory savings; 5x latency improvement**

HRTF				
Filter	MSE	Speedup	Latency	Compr.
FIR	0	1	160	1
4 Bi	3.9 (-1)	11.4	8	17.8
8 Bi	1.1 (-1)	5.7	16	9.4
16 Bi	2.5 (-2)	2.86	32	4.8
32 Bi	5.9 (-3)	1.42	64	2.5

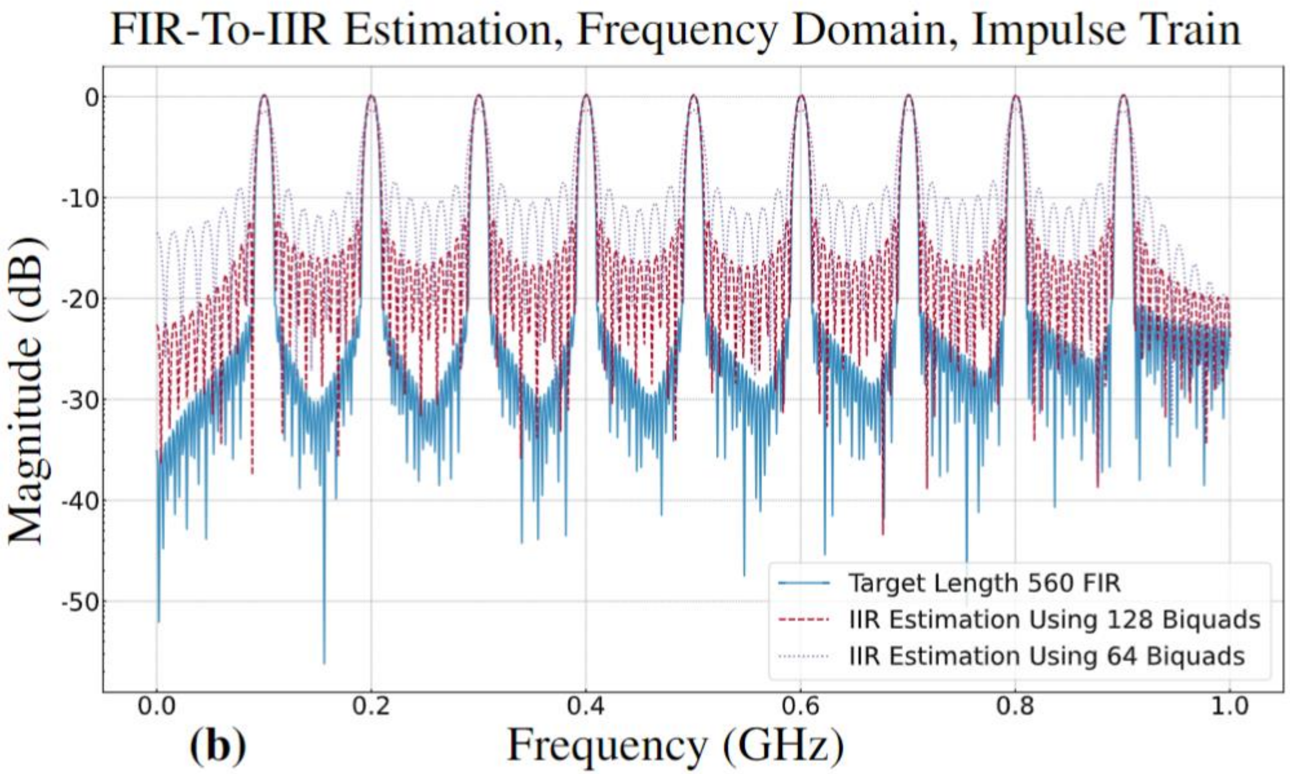
(speedup based on CPU cycle benchmarks on Cortex-M7)

Application (General Digital Filters)

- Digital FIR filters are decaying windows. To find an order N IIR estimation, we take the central length N segment time steps of the FIR.
- Similar to HRTFs, the desired spectral characteristics are preserved.



(a)



(b)

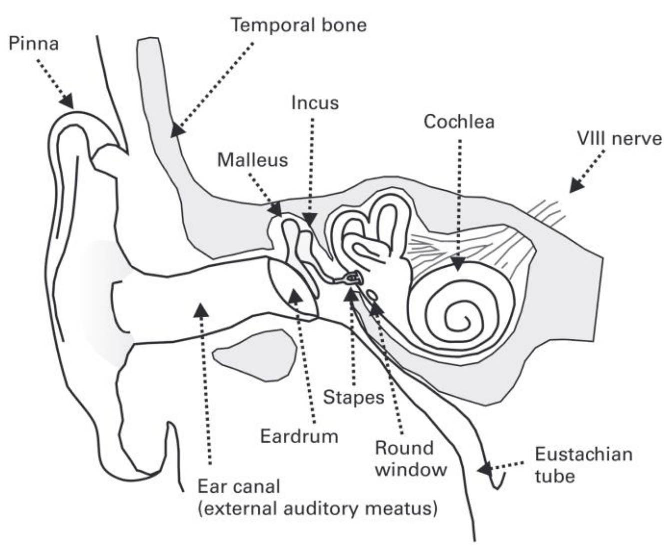
Differentiable Models of Human Hearing

Leslie Li, Dmitry N. Zotkin, Ramani Duraiswami

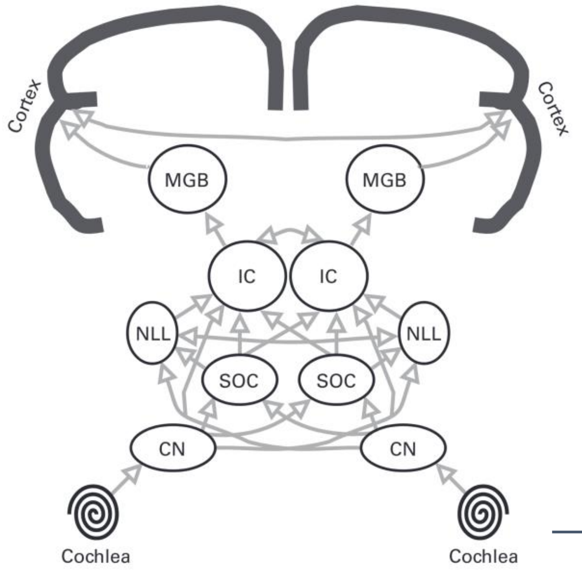
{rlli, wdz, ramanid}@umd.edu

Towards a Differentiable front end

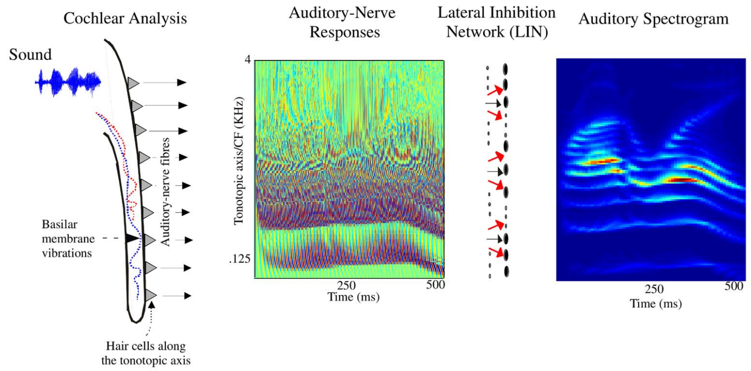
Audio →



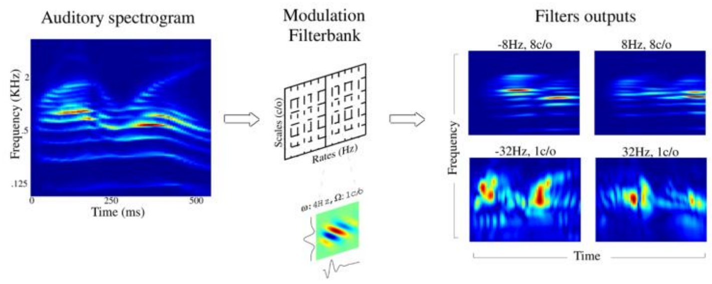
Auditory Spectrogram →



→ STRFs



Cochlear (ear)

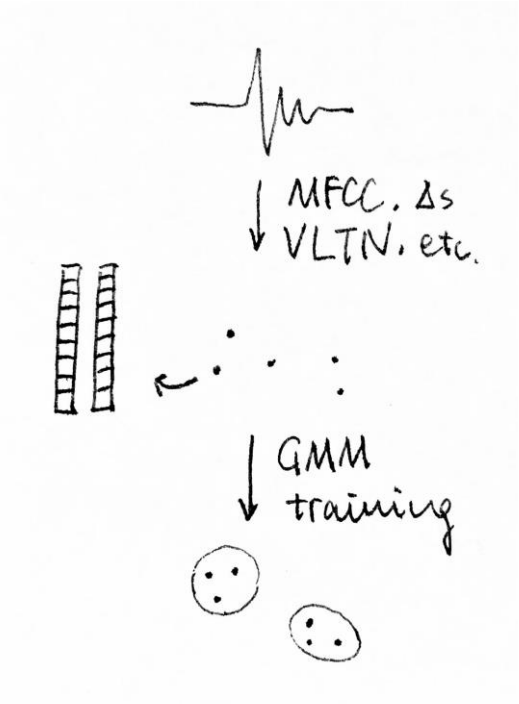


Cortical (brain)

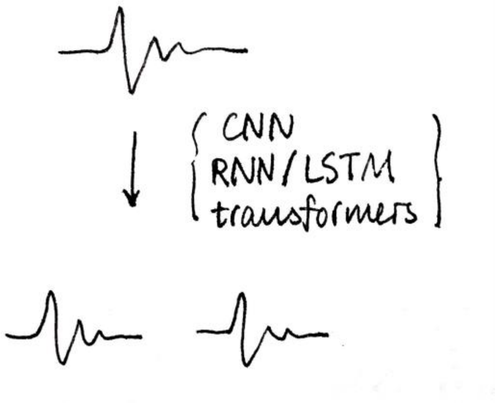
Schnupp, J., Nelken, I., & King, A. (2011). Auditory neuroscience: Making sense of sound. MIT press.

Elhilali, M. (2004). *Neural basis and computational strategies for auditory processing*. University of Maryland, College Park.

Motivation: model of perception + deep learning



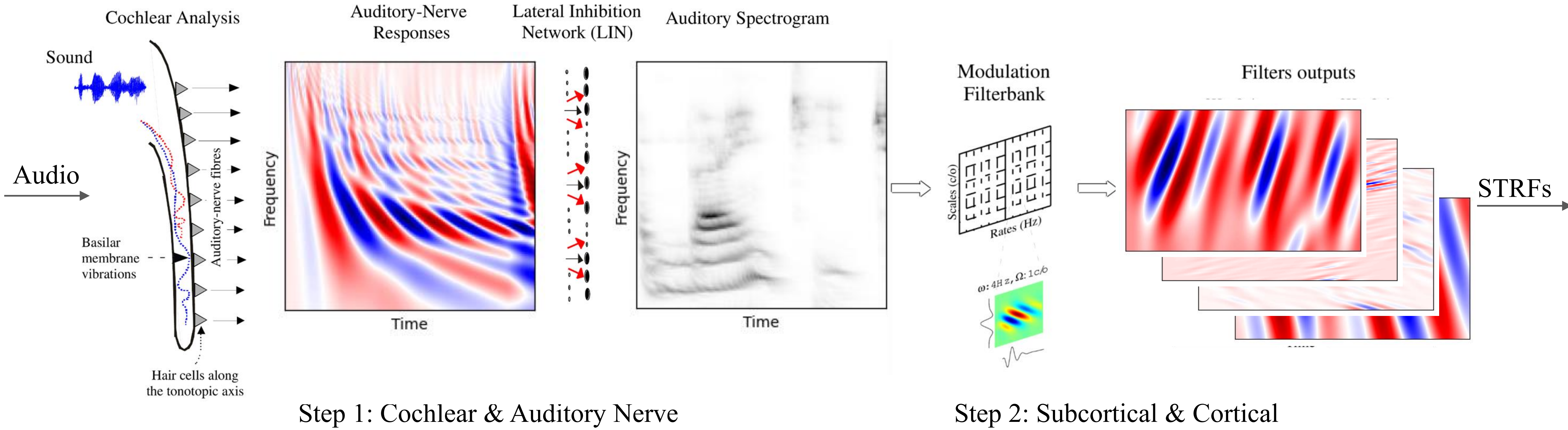
- Physically-grounded models
  - Not complex enough to handle variability in data
  - (Over-)simplifying assumptions
    - (e.g. temporal independence)
- Deep, data-driven, & end-to-end
  - Sensitive to noise and less robust in generalization
  - Large, expensive and slow



How to keep the advantages of a physical model and utilize the flexibility of deep learning?

The Model: Differentiable Auditory Processing

- Two-stage model [Chi et al., 2005]
- Implemented and made fully differentiable using JAX



Current Work

- Add these to speech processing problems