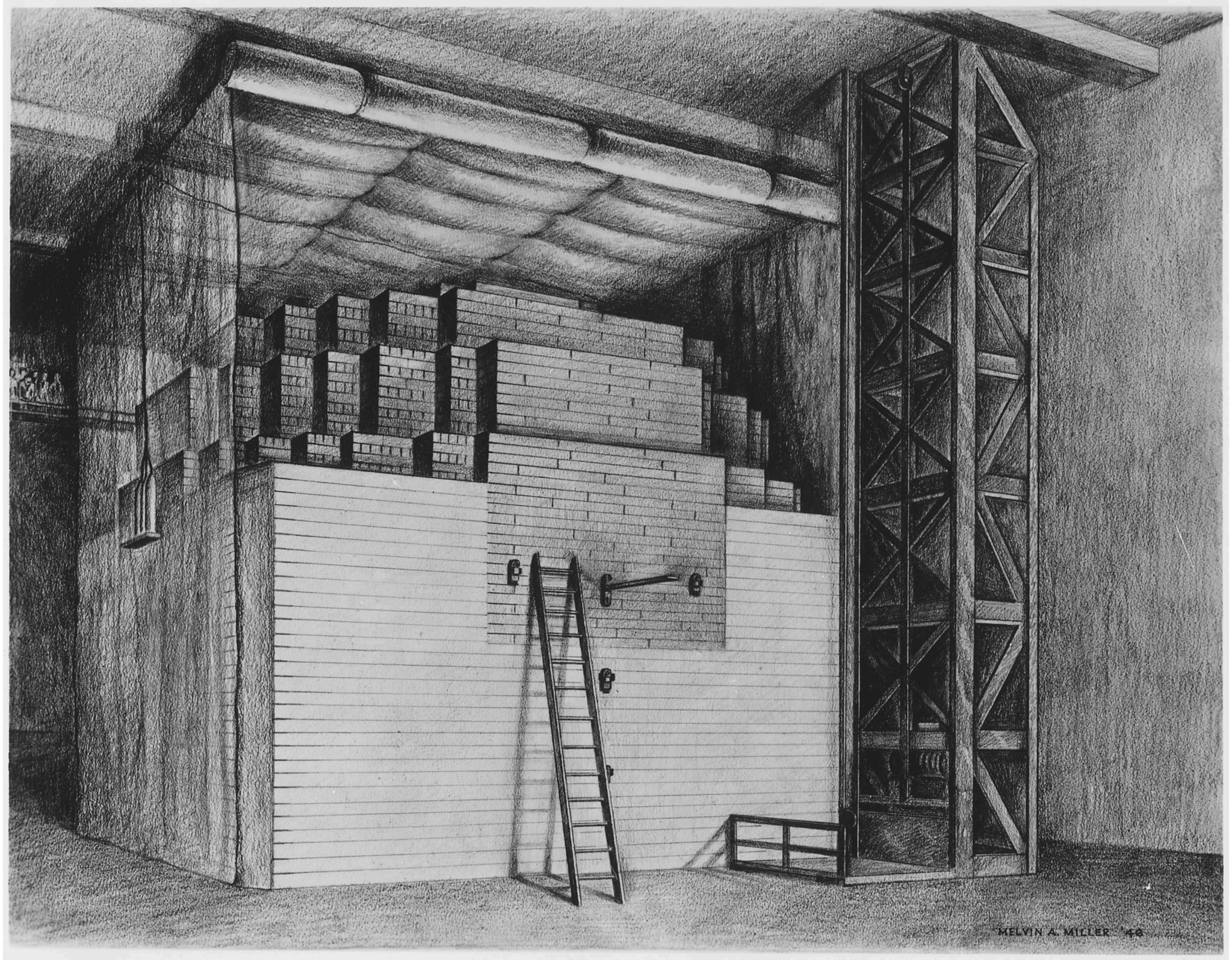
How can AI enhance measurement?

University of Chicago and Fermilab

Eric Jonas
Assistant Professor, Department of Computer Science
Committee on Computational and Applied Mathematics
Physical Sciences Division, University Of Chicago

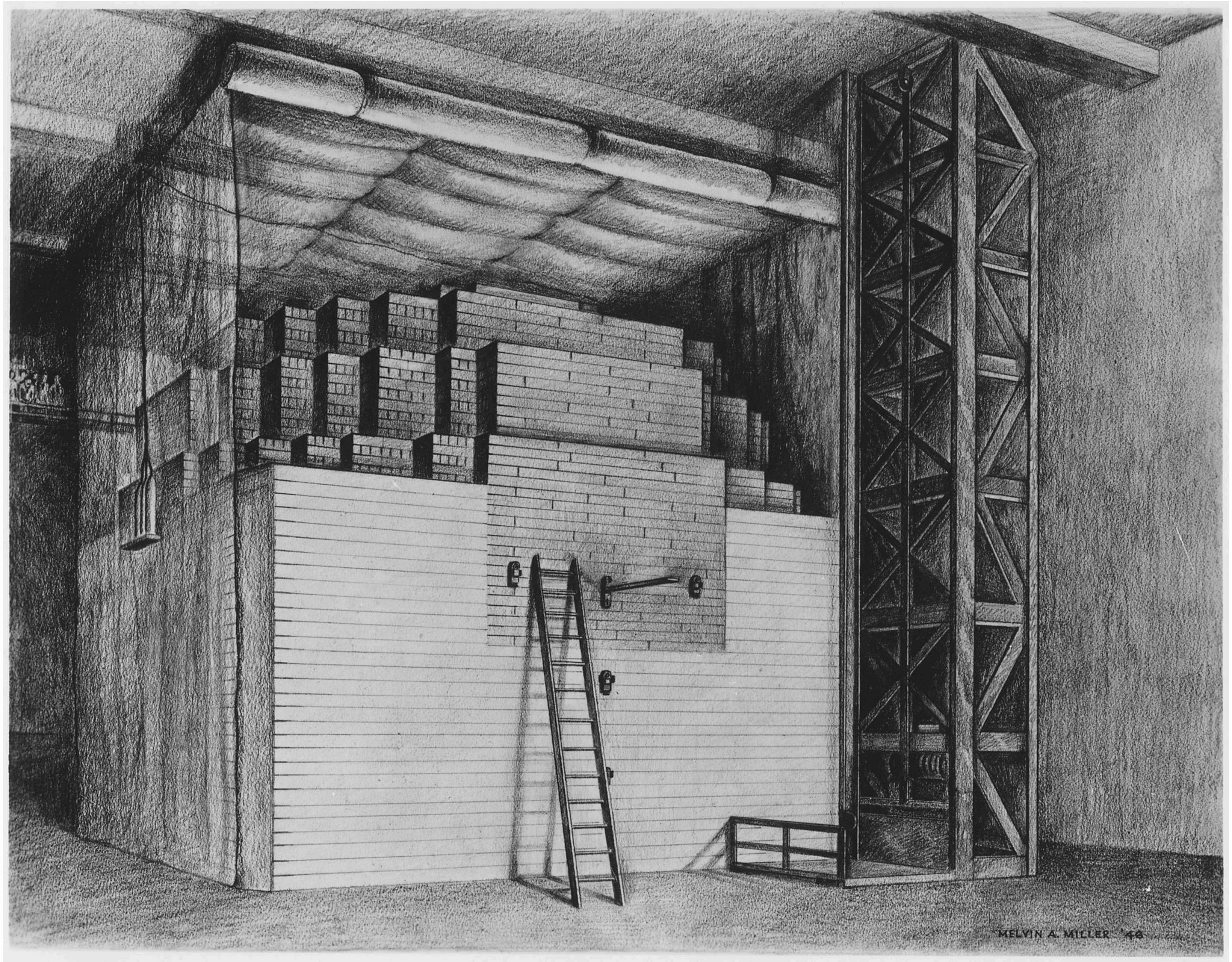
I work on AI + Scientific Measurement

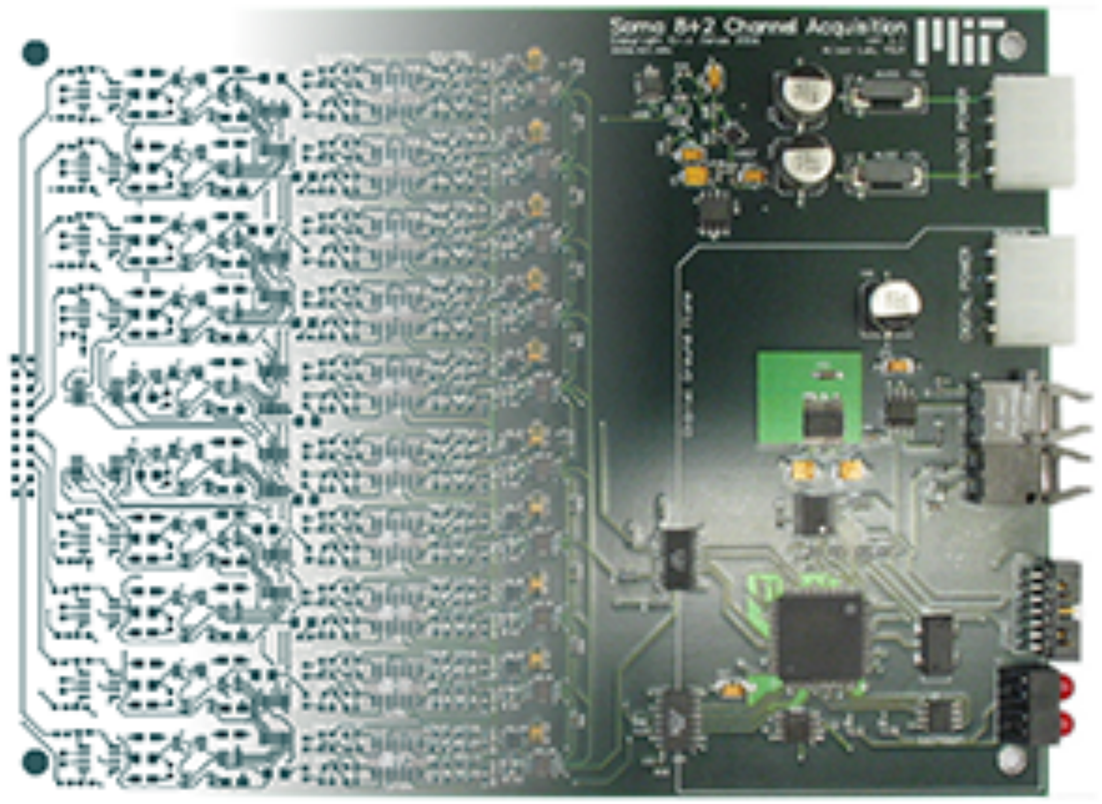
(This is the best place to be!)



I work on AI + Scientific Measurement

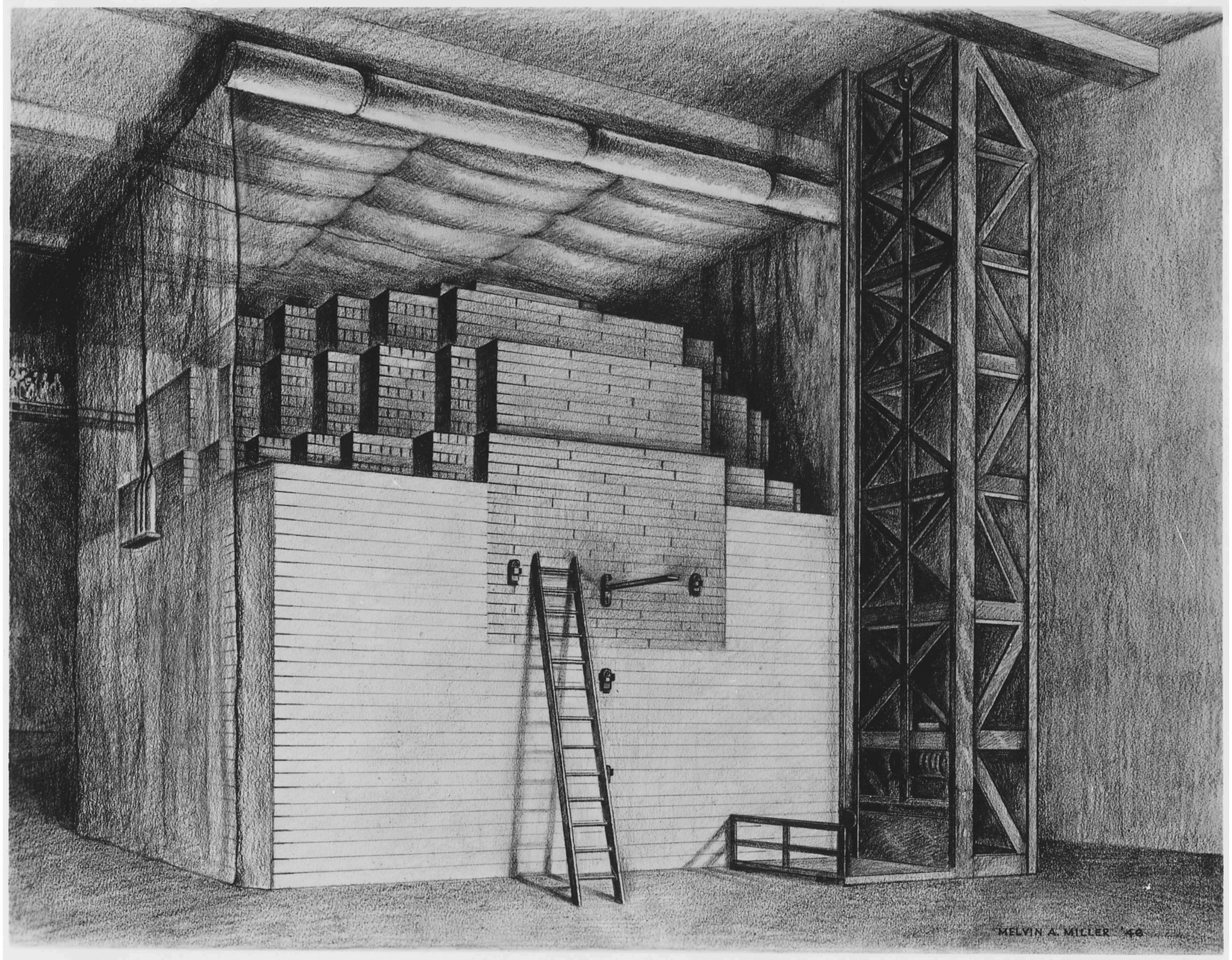
(This is the best place to be!)

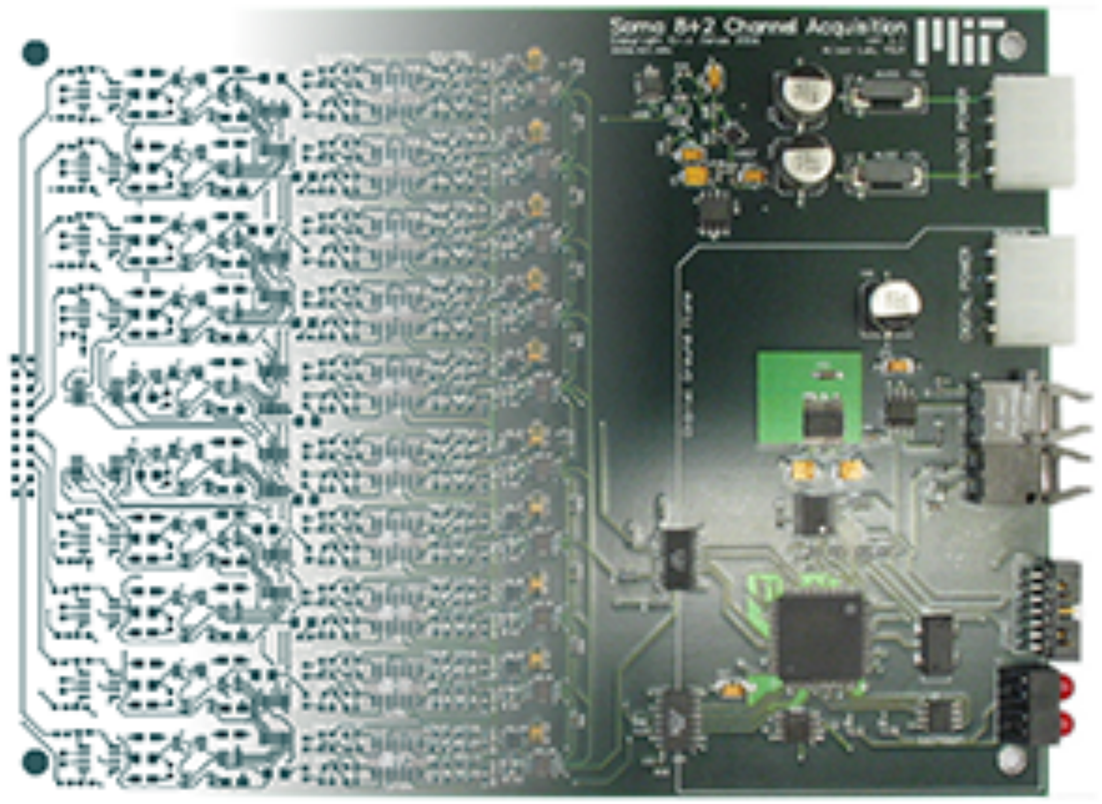


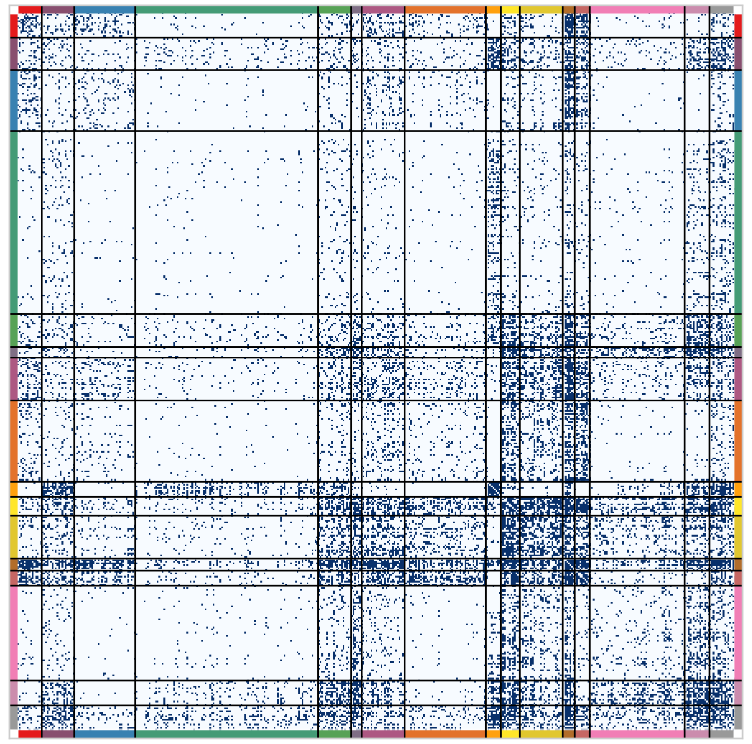


I work on AI + Scientific Measurement

(This is the best place to be!)

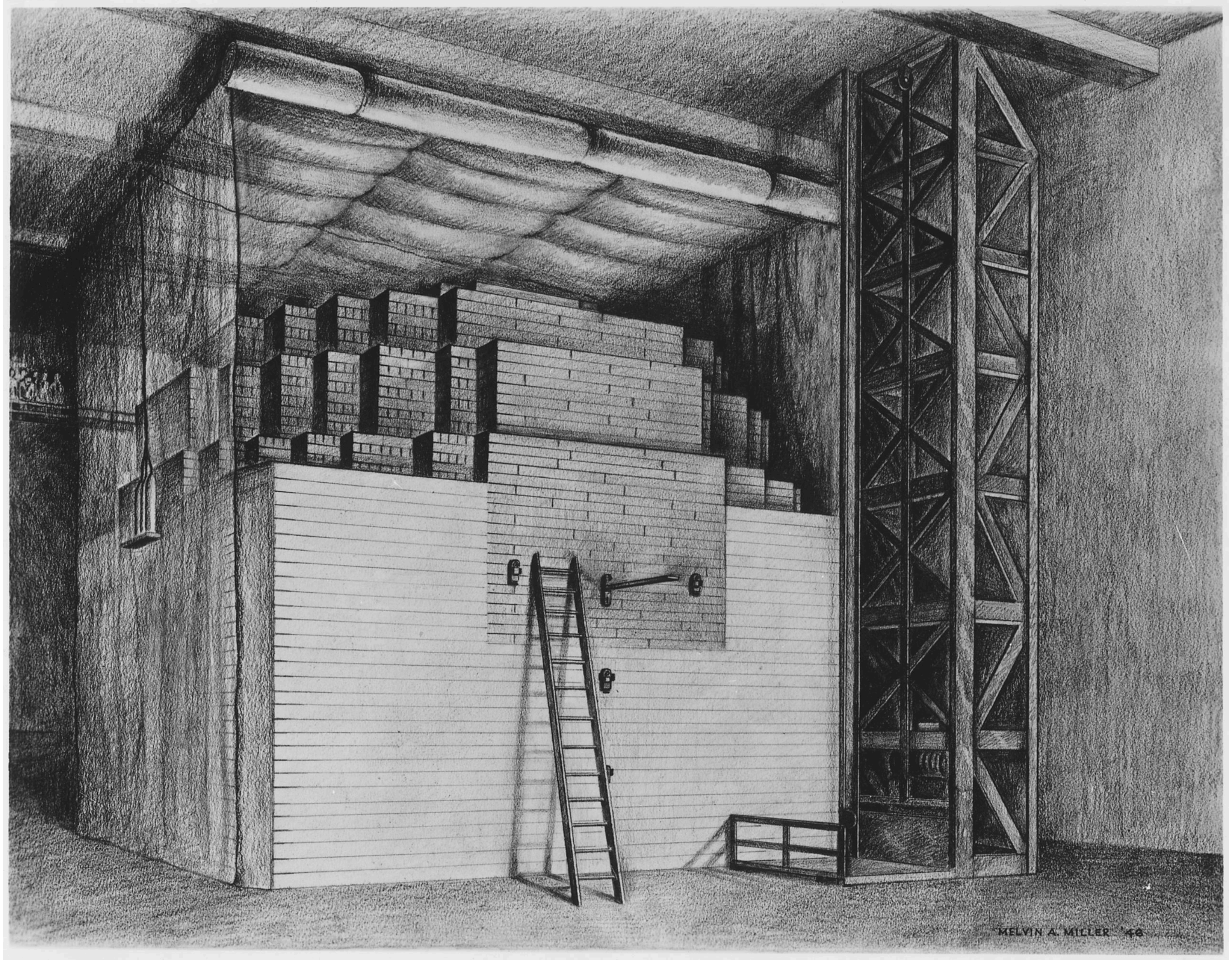


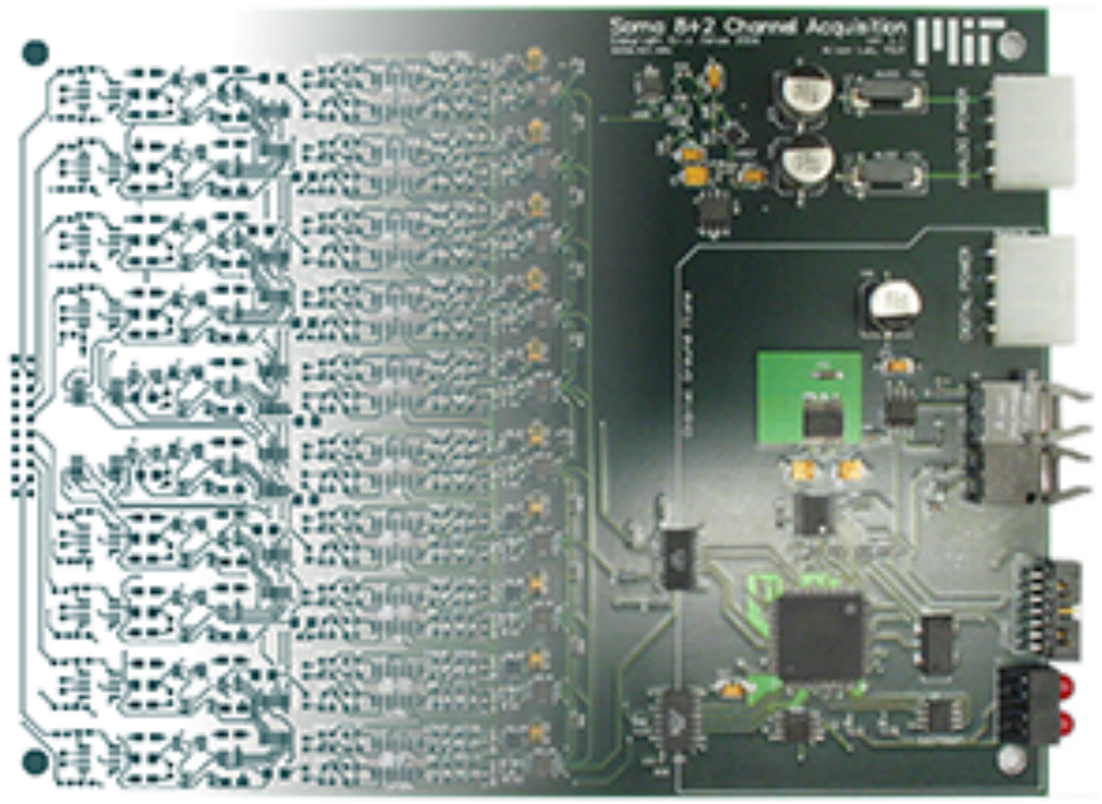


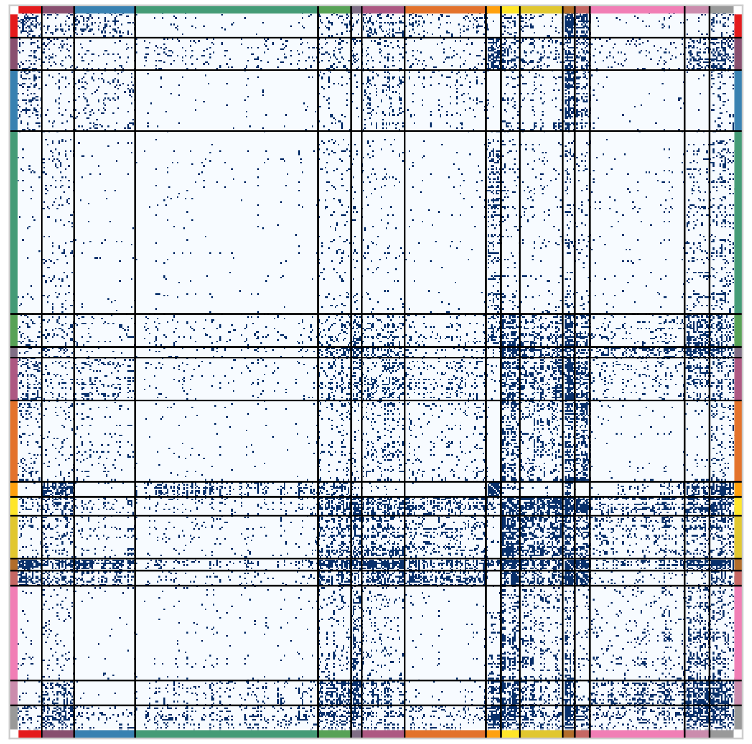


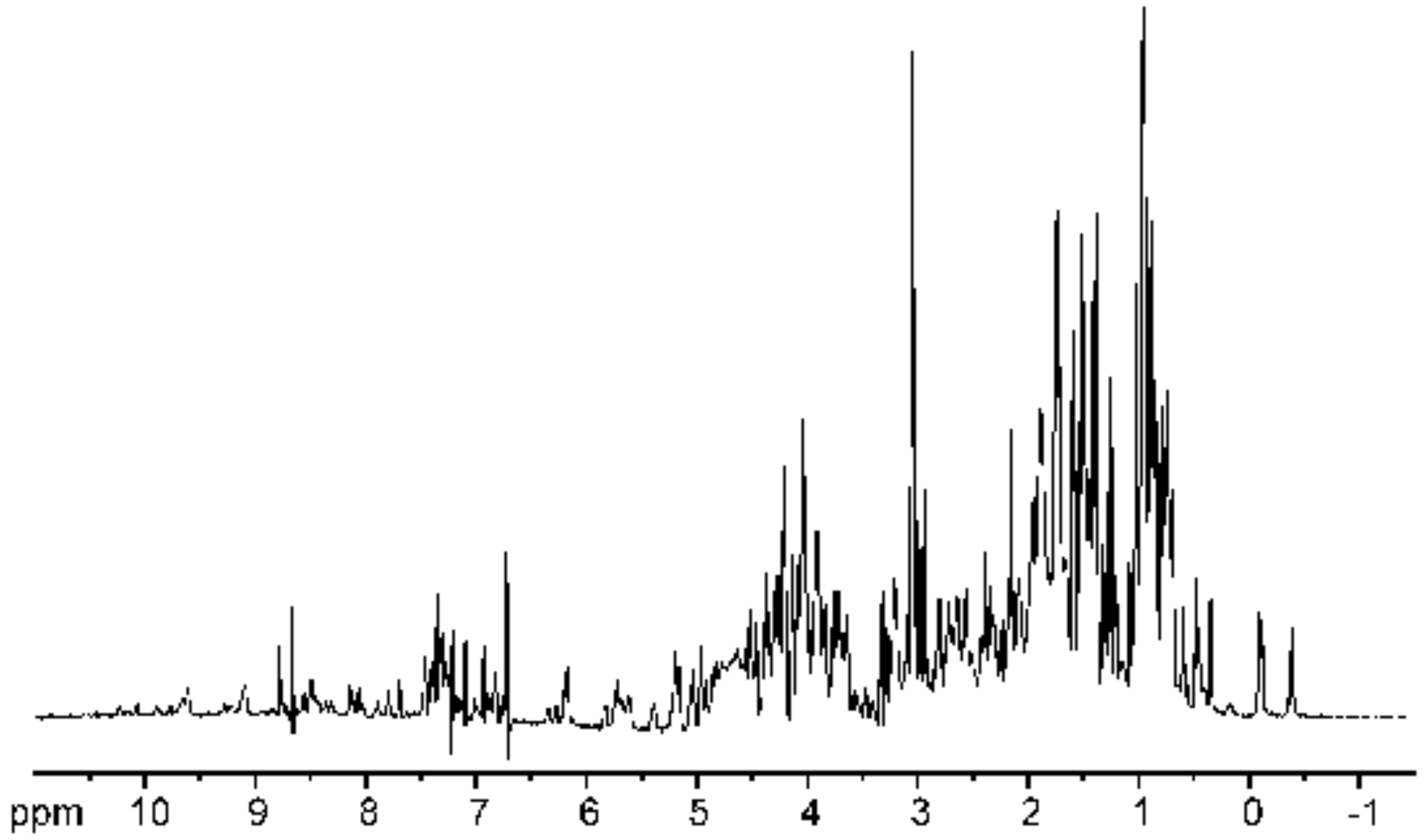
I work on AI + Scientific Measurement

(This is the best place to be!)





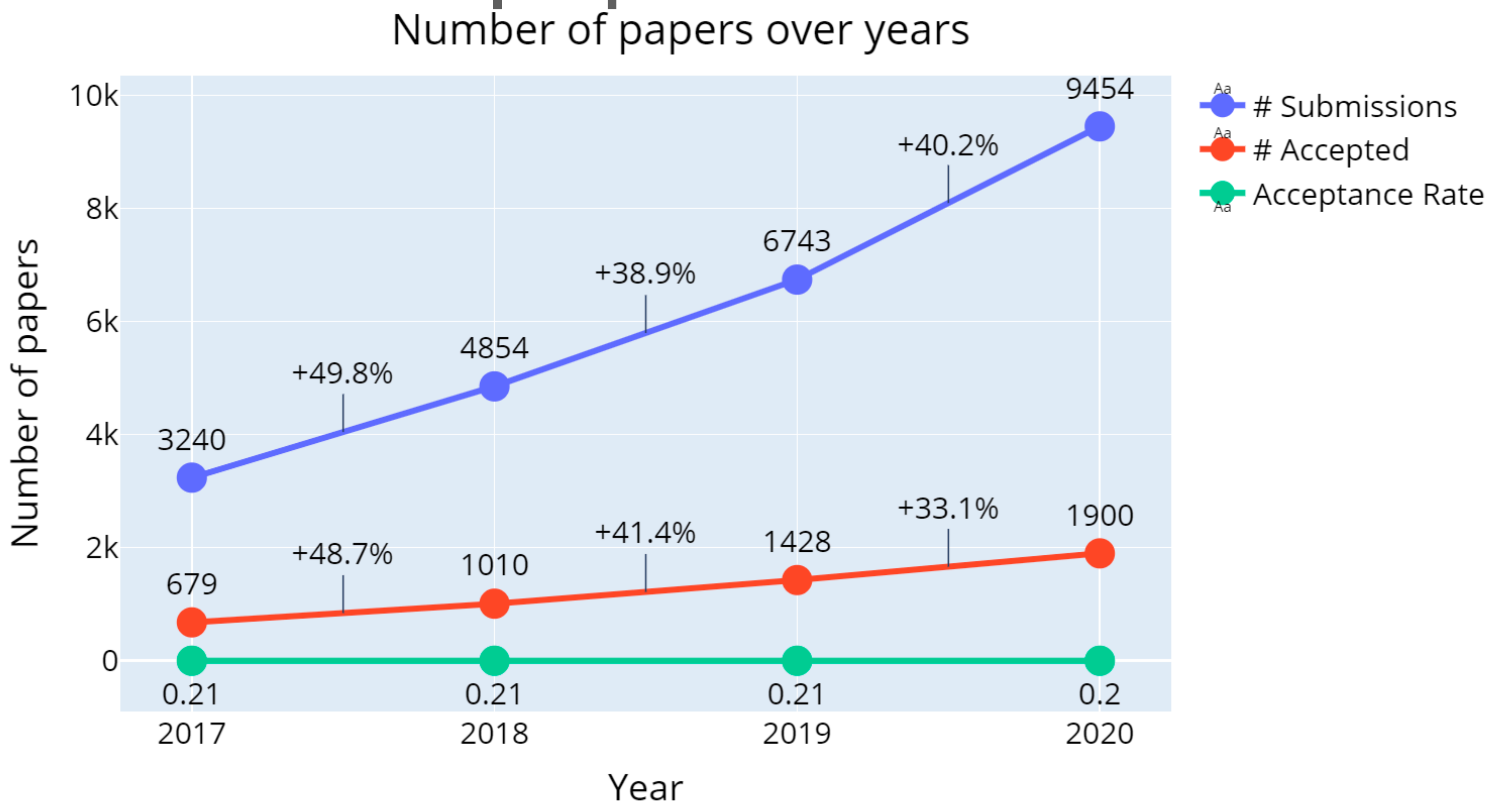




Wowza!

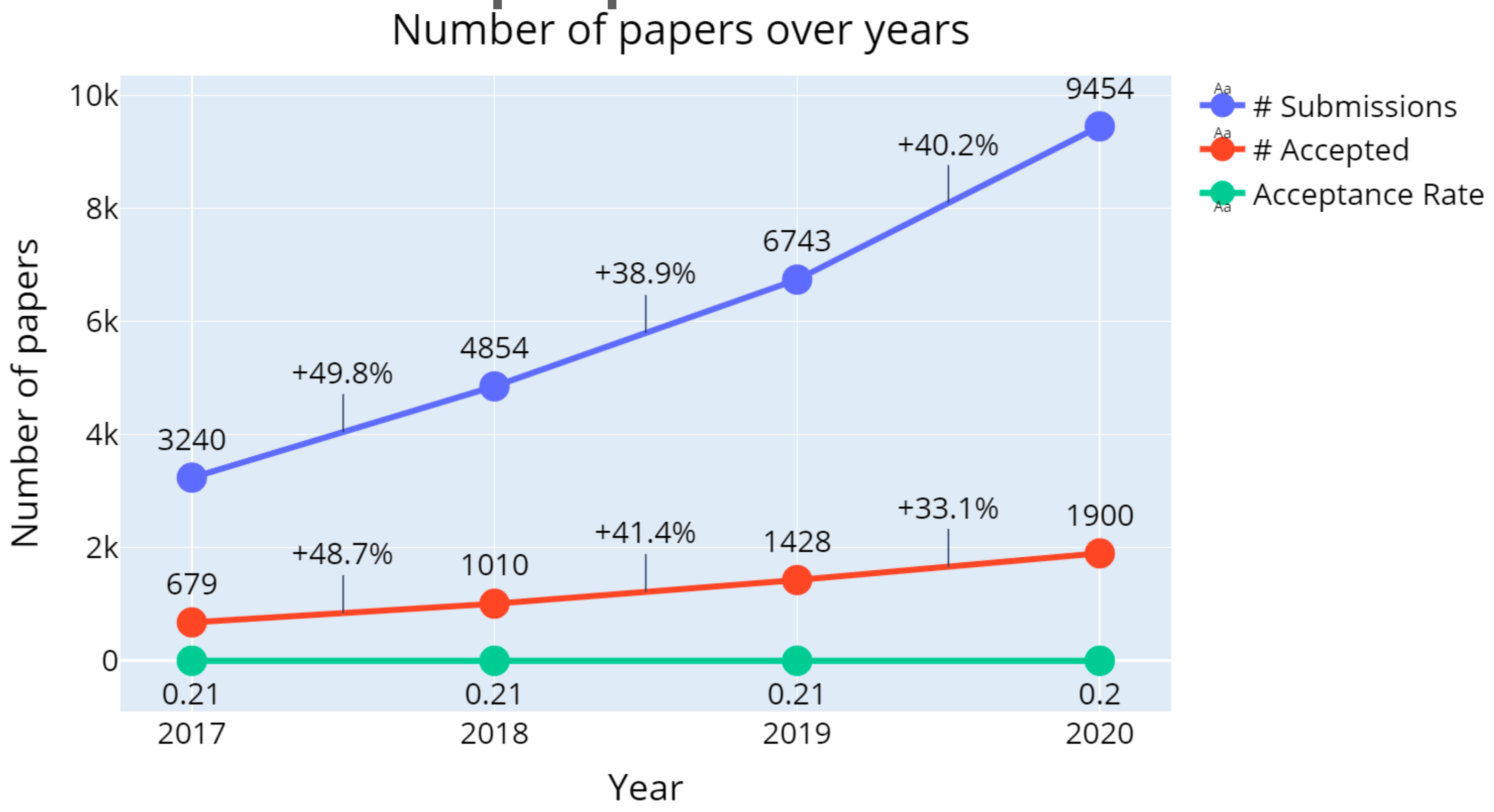
Wowza!

NeurIPS papers



Wowza!

NeurIPS papers



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > news > article

NEWS | 22 September 2022

AlphaFold developers win US$3-million Breakthrough Prize

DeepMind's system for predicting the 3D structure of proteins is among five recipients of science's most lucrative awards.

Zeeya Merali

Demis Hassabis (left) and John Jumper (right) from DeepMind developed AlphaFold, an AI that can predict the structure of proteins. Credit: Breakthrough Prize

Yikes!

Yikes!

MIT Technology Review

Sign in

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

By Will Douglas Heaven

July 30, 2021

Yikes!

MIT Technology Review

Sign in

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

By Will Douglas Heaven

July 30, 2021

Google medical researchers humbled when AI screening tool falls short in real-life testing

Devin Coldewey @techcrunch / 4:03 PM CDT • April 27, 2020

Comment

Yikes!

MIT Technology Review

Sign in

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

By Will Douglas Heaven

July 30, 2021

Google medical researchers humbled when AI screening tool falls short in real-life testing

Devin Coldewey @techcrunch / 4:03 PM CDT • April 27, 2020 Comment

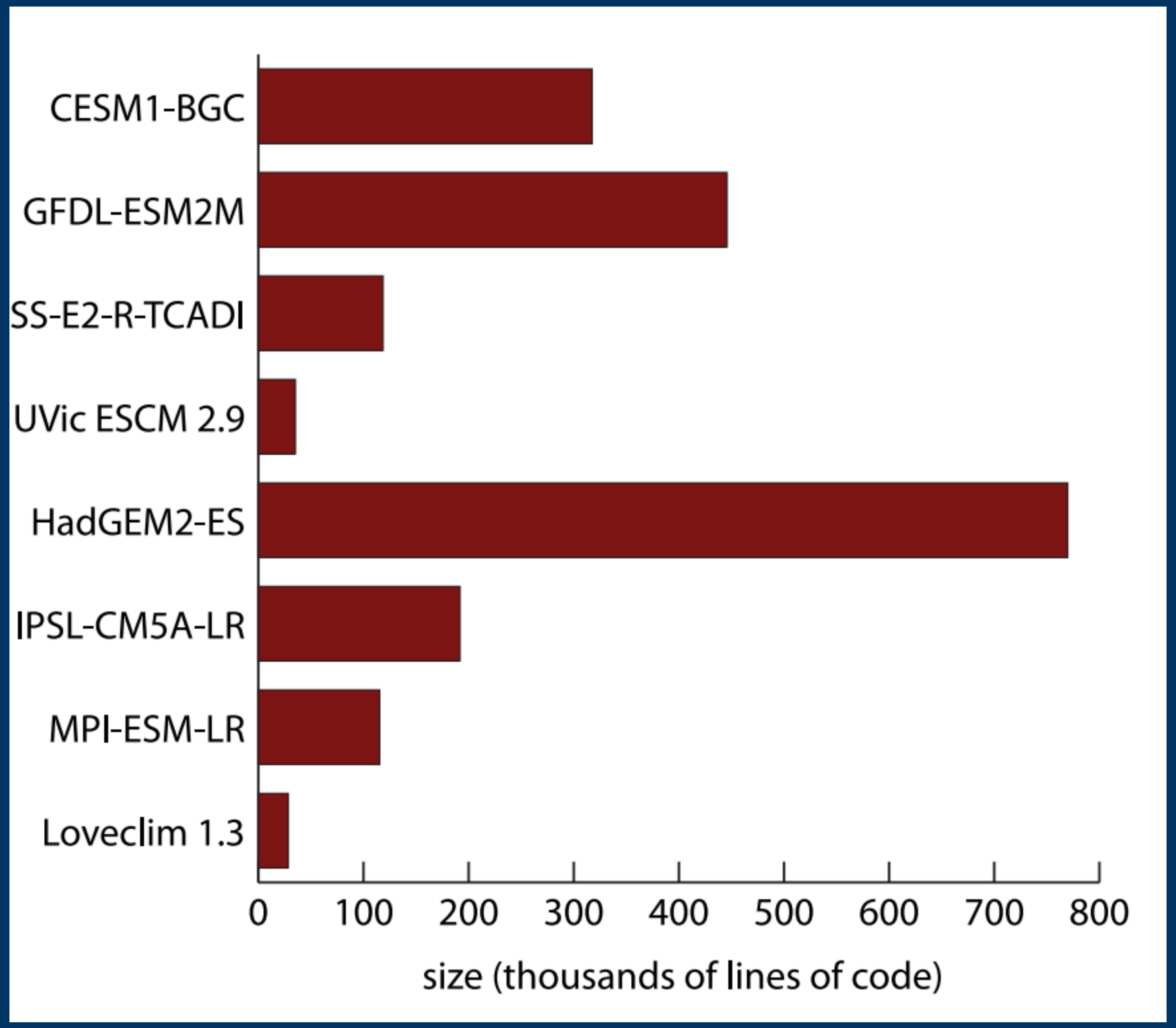
Leakage and the Reproducibility Crisis in ML-based Science

Sayash Kapoor [1] Arvind Narayanan [1]

Field	Paper	Number of papers reviewed	Number of papers with pitfalls	[L1.1] No test set	[L1.2] Pre-proc. on train-test	[L1.3] Feature sel. on train-test	[L1.4] Duplicates	[L2] Illegitimate features	[L3.1] Temporal leakage	[L3.2] Non-ind. b/w train-test	[L3.3] Sampling bias	Comput. reproducibility issues	Data quality issues	Metric choice issues	Standard dataset used?
Medicine	Bouwmeester et al. (2012)	71	27	○									○		
Neuroimaging	Whelan & Garavan (2014)	–	14	○		○									
Autism Diagnostics	Bone et al. (2015)	–	3				○				○		○	○	○
Bioinformatics	Blagus & Lusa (2015)	–	6		○										
Nutrition Research	Ivanescu et al. (2016)	–	4	○									○	○	
Software Eng.	Tu et al. (2018)	58	11						○			○	○		○
Toxicology	Alves et al. (2019)	–	1				○					○	○		
Satellite Imaging	Nalepa et al. (2019)	17	17							○			○		○
Tractography	Poulin et al. (2019)	4	2	○								○	○	○	○
Clinical Epidem.	Christodoulou et al. (2019)	71	48			○							○		
Brain-computer Int.	Nakanishi et al. (2020)	–	1	○											○
Histopathology	Oner et al. (2020)	–	1							○					
Neuropsychiatry	Poldrack et al. (2020)	100	53	○	○								○	○	
Medicine	Vandewiele et al. (2021)	24	21			○				○	○	○	○		○
Radiology	Roberts et al. (2021)	62	62	○			○				○	○			○
IT Operations	Lyu et al. (2021)	9	3						○						○
Medicine	Filho et al. (2021)	–	1					○							
Neuropsychiatry	Shim et al. (2021)	–	1			○						○			
Genomics	Barnett et al. (2022)	41	23			○							○		
Computer Security	Arp et al. (2022)	30	30	○	○	○		○		○	○		○	○	

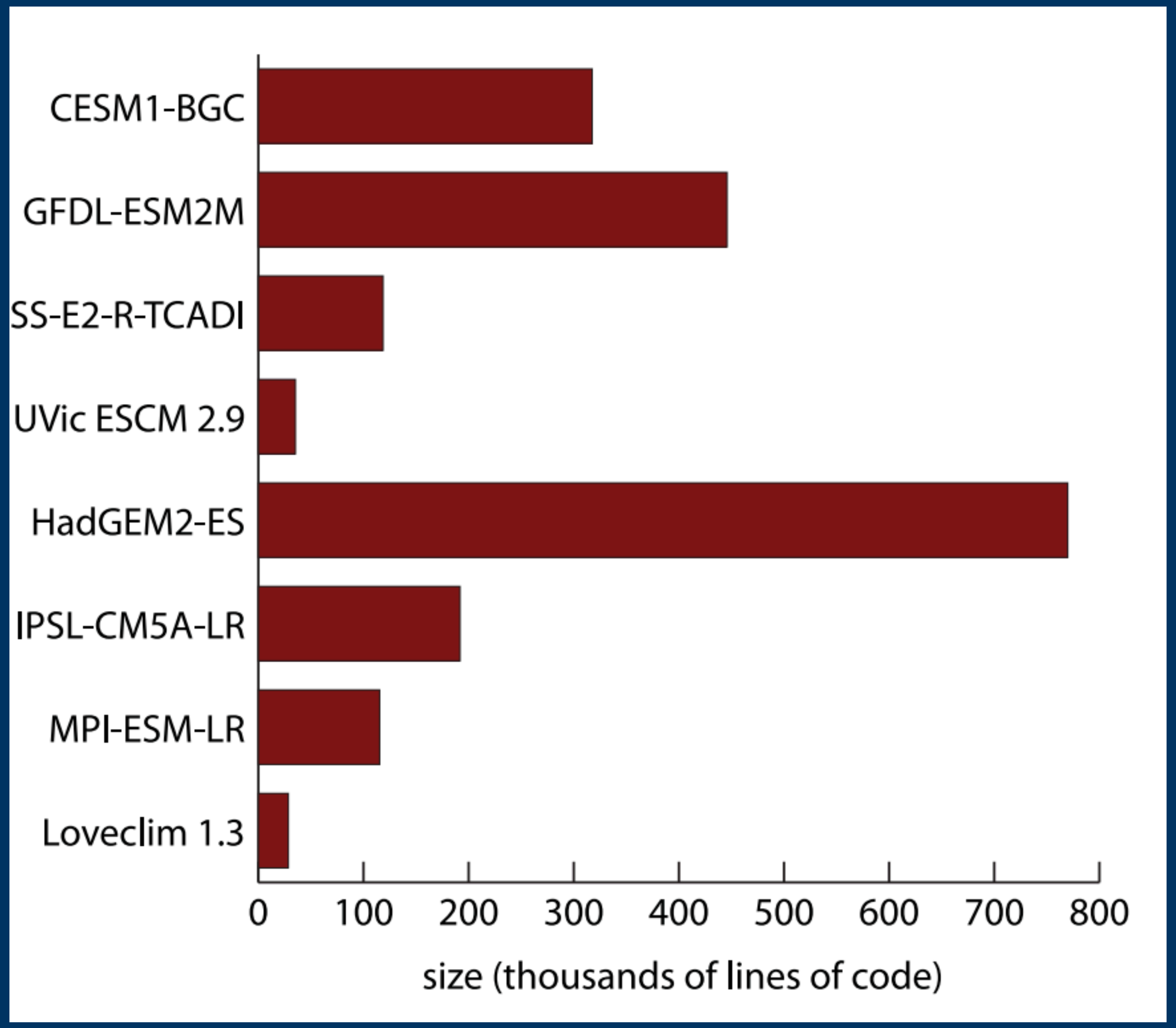
https://reproducible.cs.princeton.edu/

What makes
AI for Experiments
special?



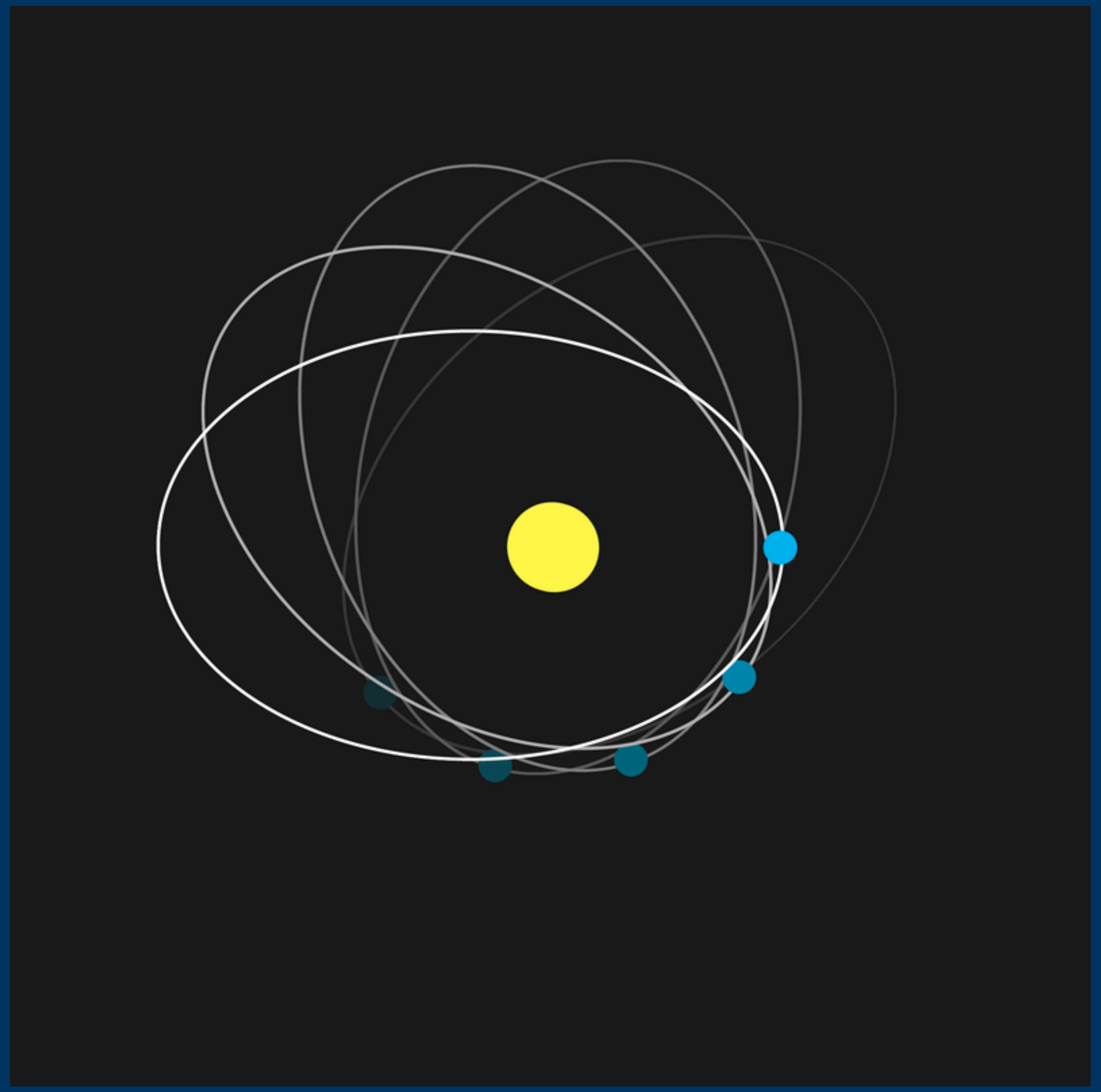
We take our models seriously

Alexander, Kaitlin, and Stephen M. Easterbrook. "The software architecture of climate models: a graphical comparison of CMIP5 and EMICAR5 configurations." Geoscientific Model Development 8.4 (2015): 1221-1232.



We take our models seriously

Alexander, Kaitlin, and Stephen M. Easterbrook. "The software architecture of climate models: a graphical comparison of CMIP5 and EMICAR5 configurations." Geoscientific Model Development 8.4 (2015): 1221-1232.



Distribution shift is the point

Anomalous rate of precession of the perihelion of Mercury
https://en.wikipedia.org/wiki/Tests_of_general_relativity

Need to think carefully about baselines

Simple random search provides a competitive approach to reinforcement learning

Horia Mania   Aurelia Guy   Benjamin Recht

"A common belief in model-free reinforcement learning is that methods based on random search in the parameter space of policies exhibit significantly worse sample complexity than those that explore the space of actions. **We dispel such beliefs by introducing a random search method for training static, linear policies for continuous control problems, matching state-of-the-art sample efficiency on the benchmark MuJoCo locomotion tasks.**"

arXiv:1803.07055v1 March 2018

An opinionated

taxonomy

For AI + Experiments

An opinionated

taxonomy

For AI + Experiments

Inverse Problems

How can AI help us *understand* the data that we get from existing systems

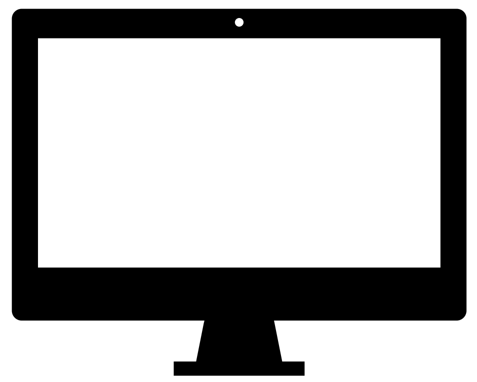
An opinionated

taxonomy

For AI + Experiments

Inverse Problems

How can AI help us *understand* the data that we get from existing systems



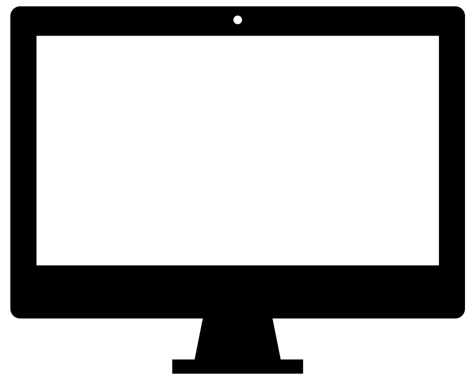
Computational Measurement

How can we design *new measurement systems* to be more interpretable by AI?

An opinionated
taxonomy
For AI + Experiments

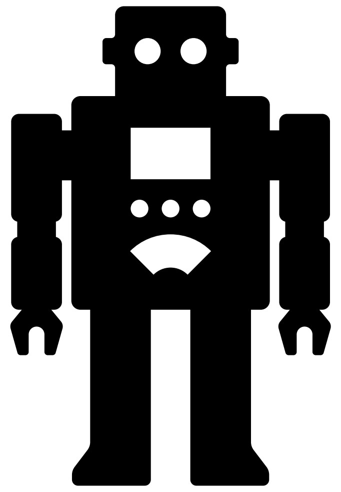
Inverse Problems

How can AI help us *understand* the data that we get from existing systems



Computational Measurement

How can we design *new measurement systems* to be more interpretable by AI?



Active Learning

How can AI *guide experimentation* and measurement?

An opinionated

taxonomy

For AI + Experiments

Inverse Problems

How can AI help us *understand* the data that we get from existing systems

Computational Measurement

How can we design *new measurement systems* to be more interpretable by AI?

Active Learning

How can AI *guide experimentation* and measurement?

Our goal: Create new collaborations, identify under-explored areas, set the stage for next round of funding

AI for Inverse Problems

How can AI help us *understand* the data
that we get from existing systems

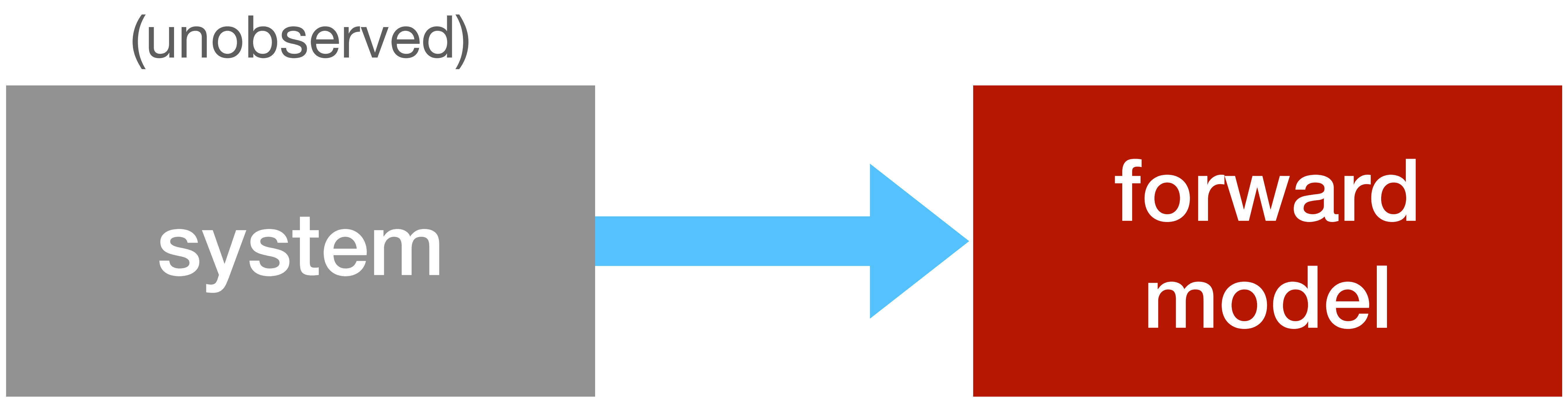
Inverse Problems and Machine Learning

Inverse Problems and Machine Learning

(unobserved)

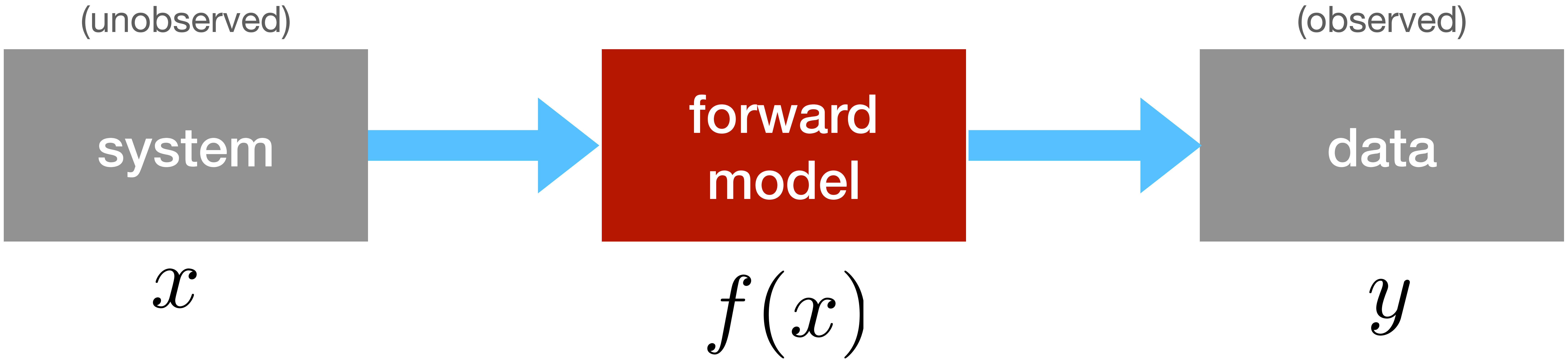
system

Inverse Problems and Machine Learning



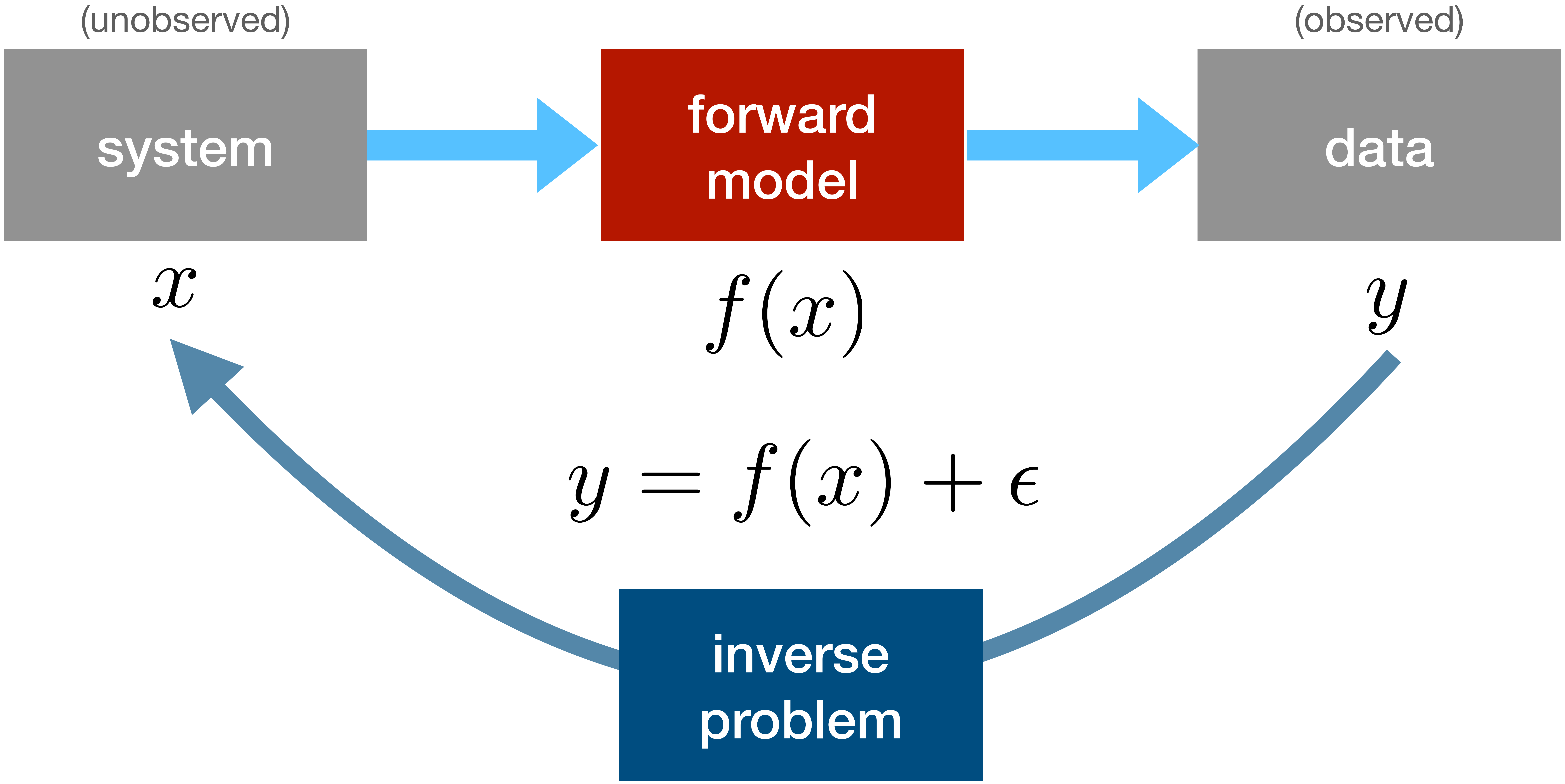
Inverse Problems and Machine Learning

Inverse Problems and Machine Learning

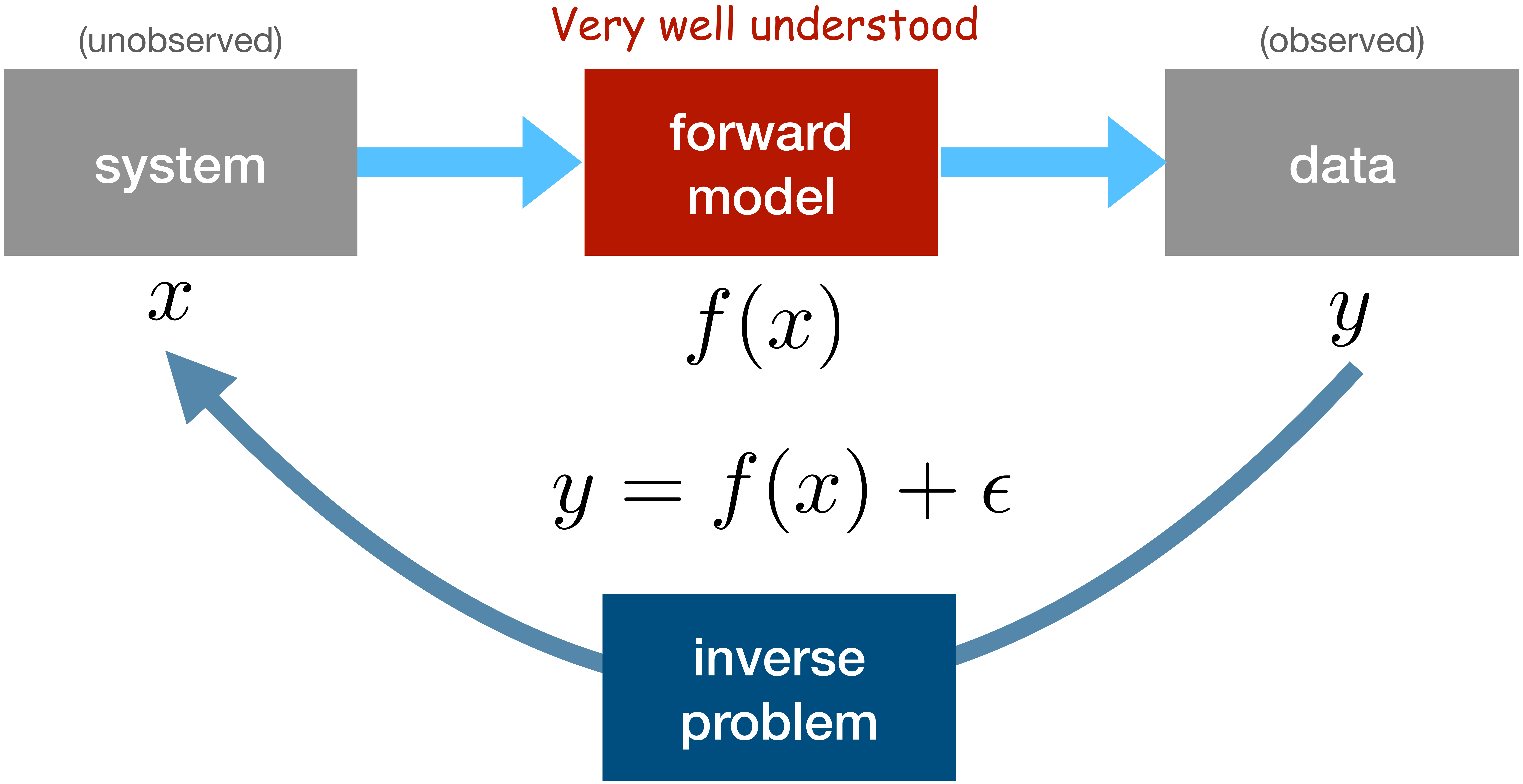


$$y = f(x) + \epsilon$$

Inverse Problems and Machine Learning



Inverse Problems and Machine Learning

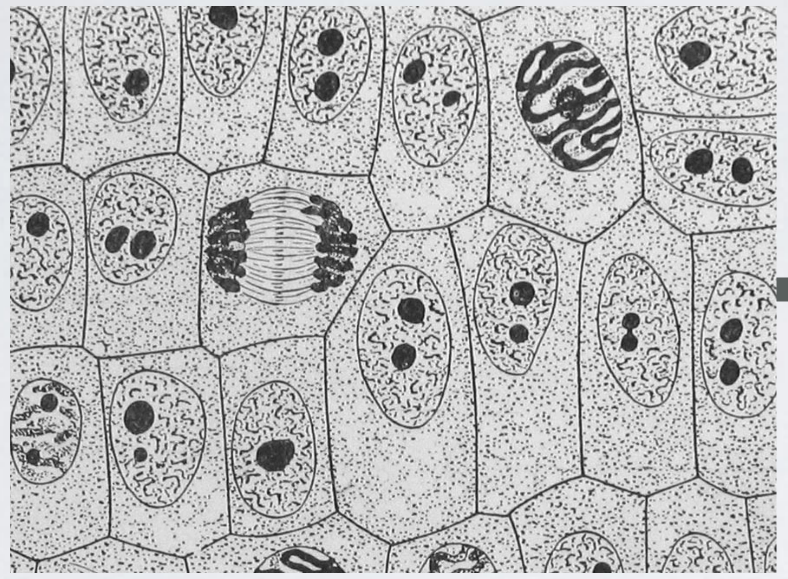


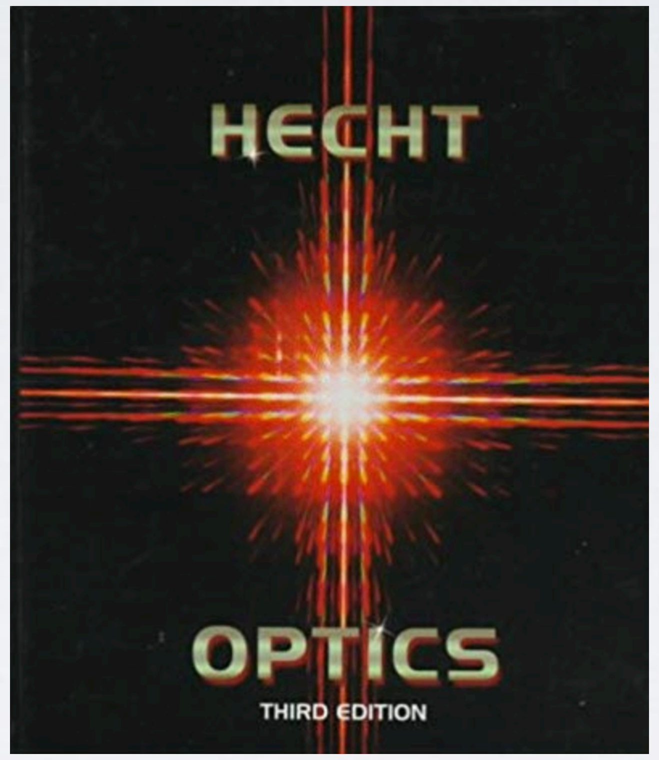
EXAMPLE PROBLEMS

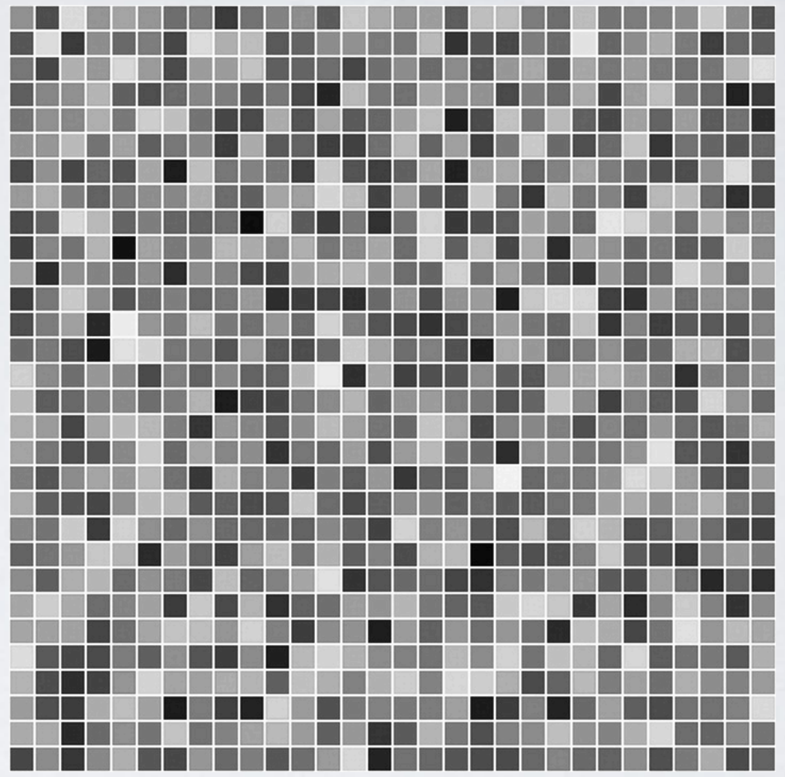
Microscopy

EXAMPLE PROBLEMS

Microscopy

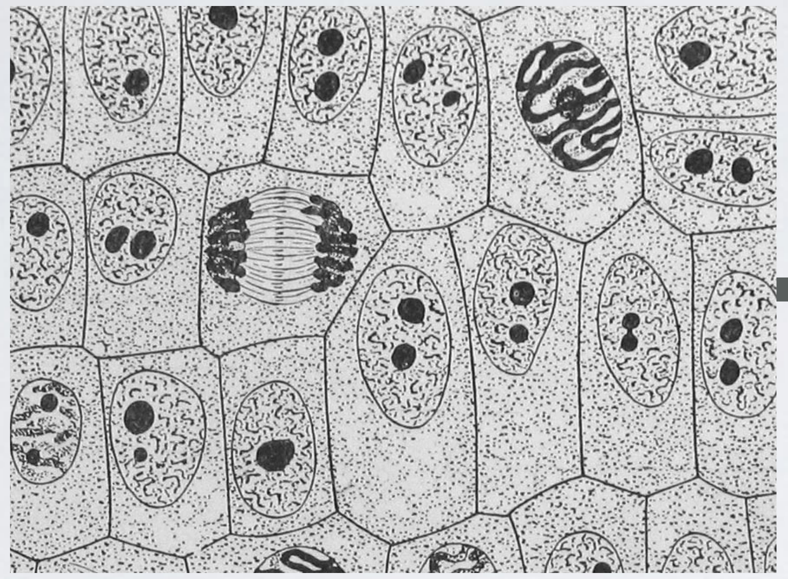


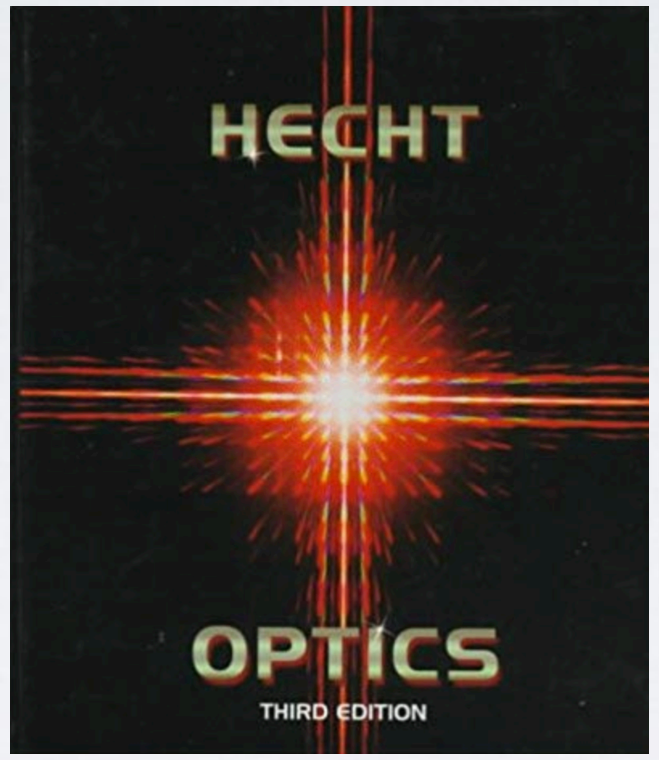


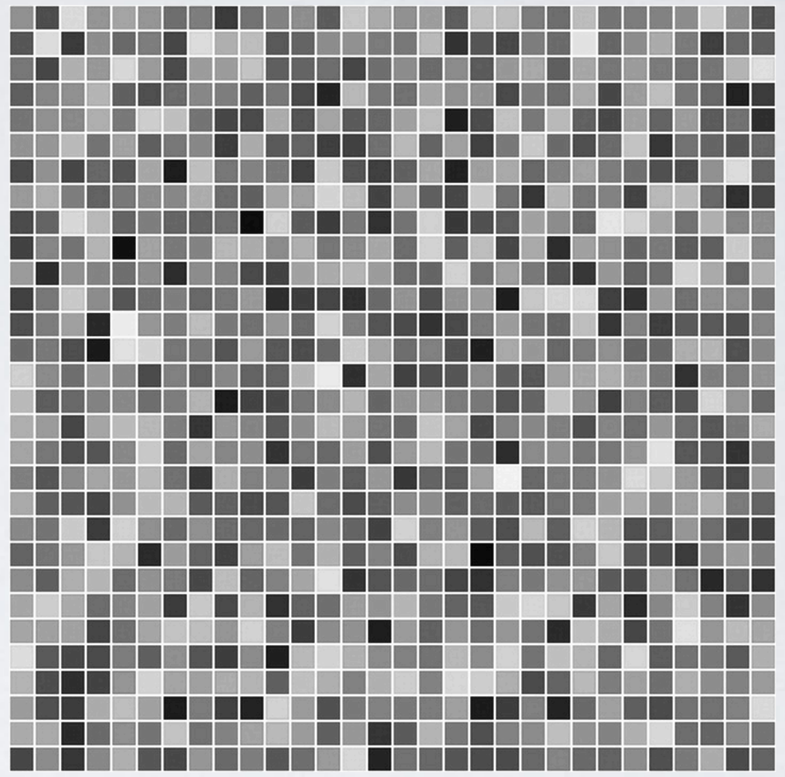


EXAMPLE PROBLEMS

Microscopy

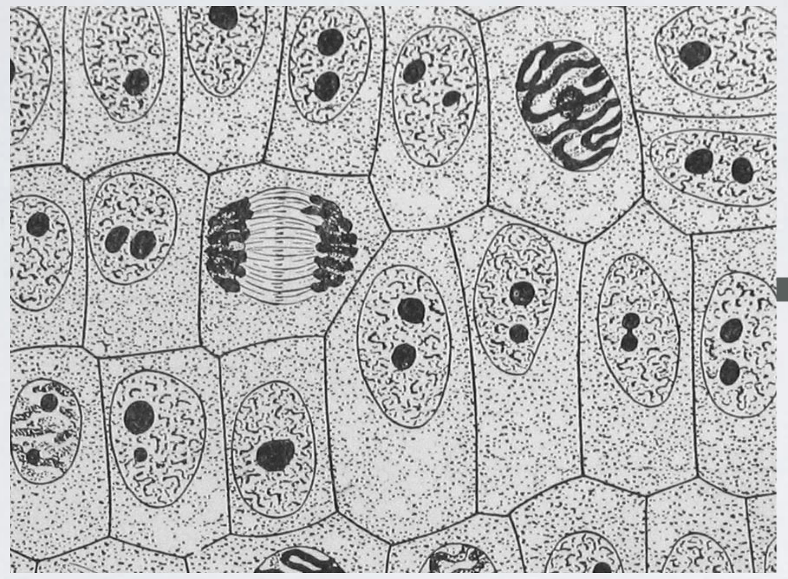


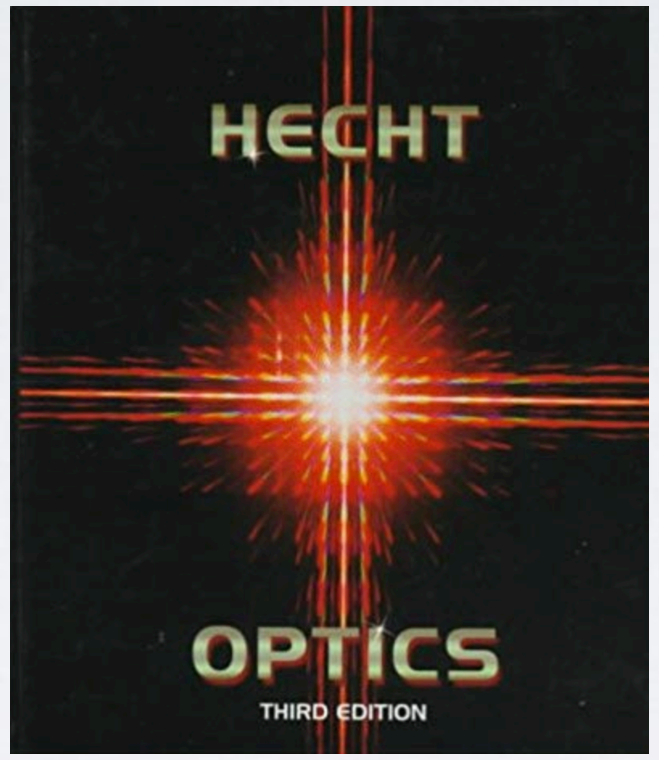


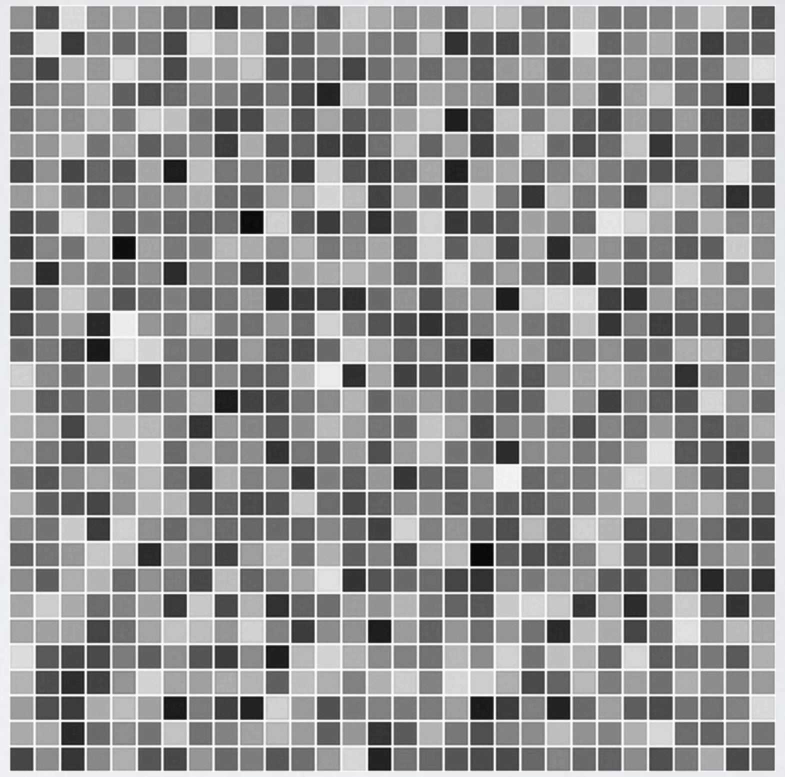


EXAMPLE PROBLEMS

Microscopy



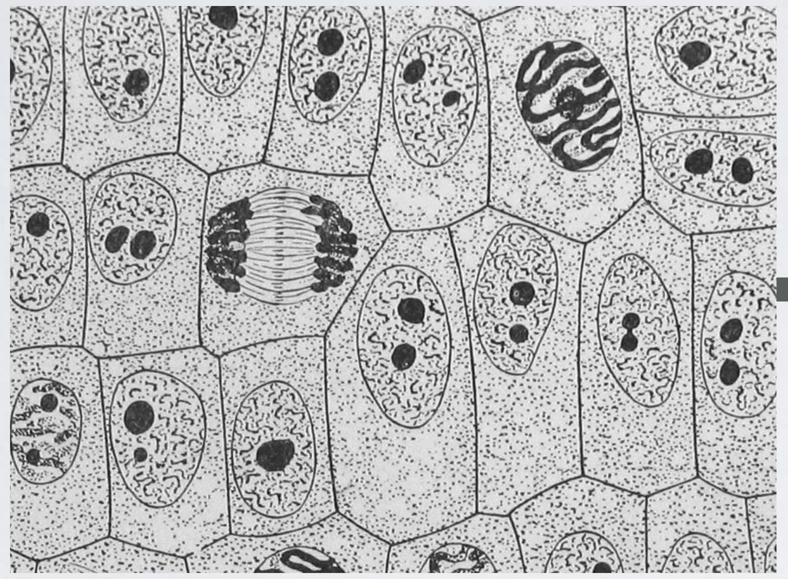


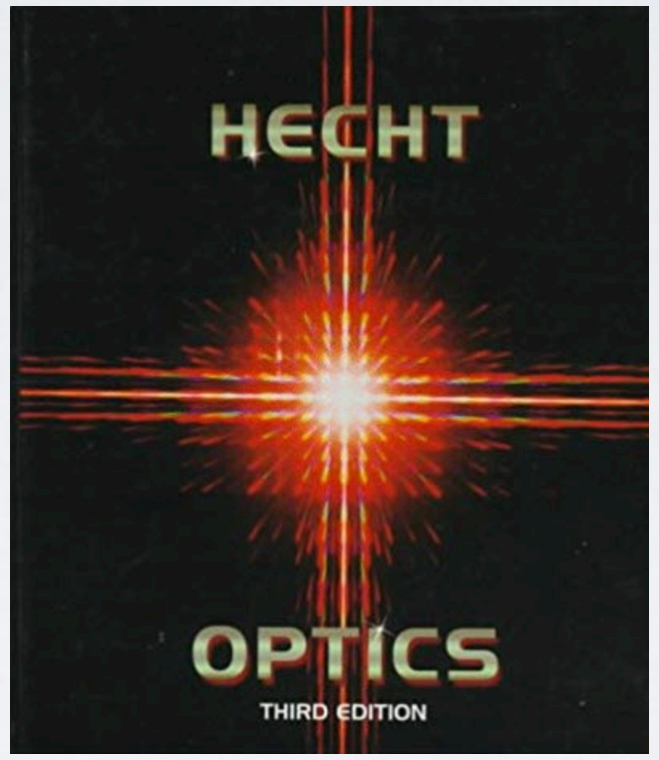


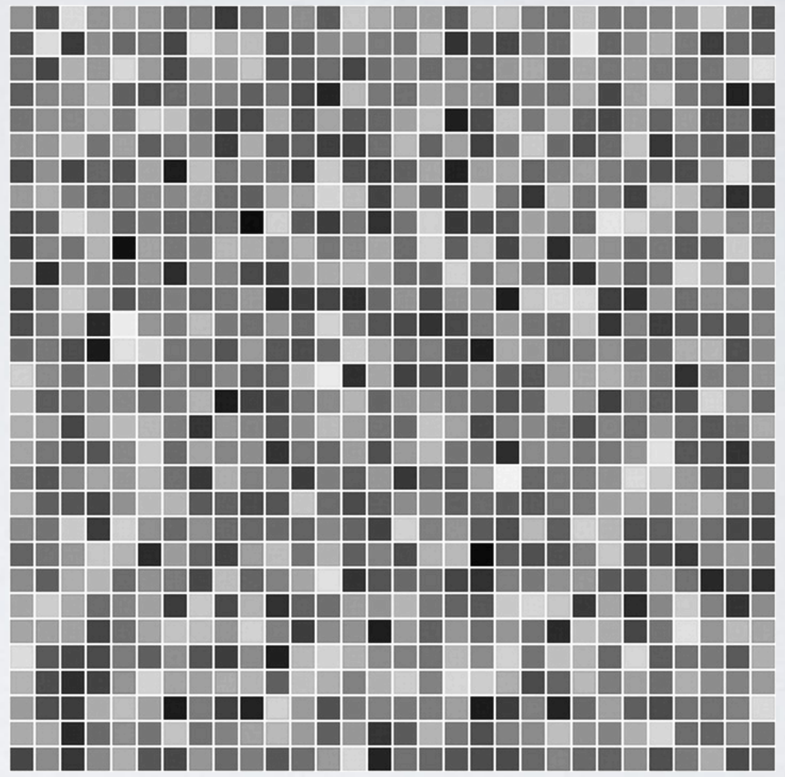
Magnetic Resonance Imaging

EXAMPLE PROBLEMS

Microscopy





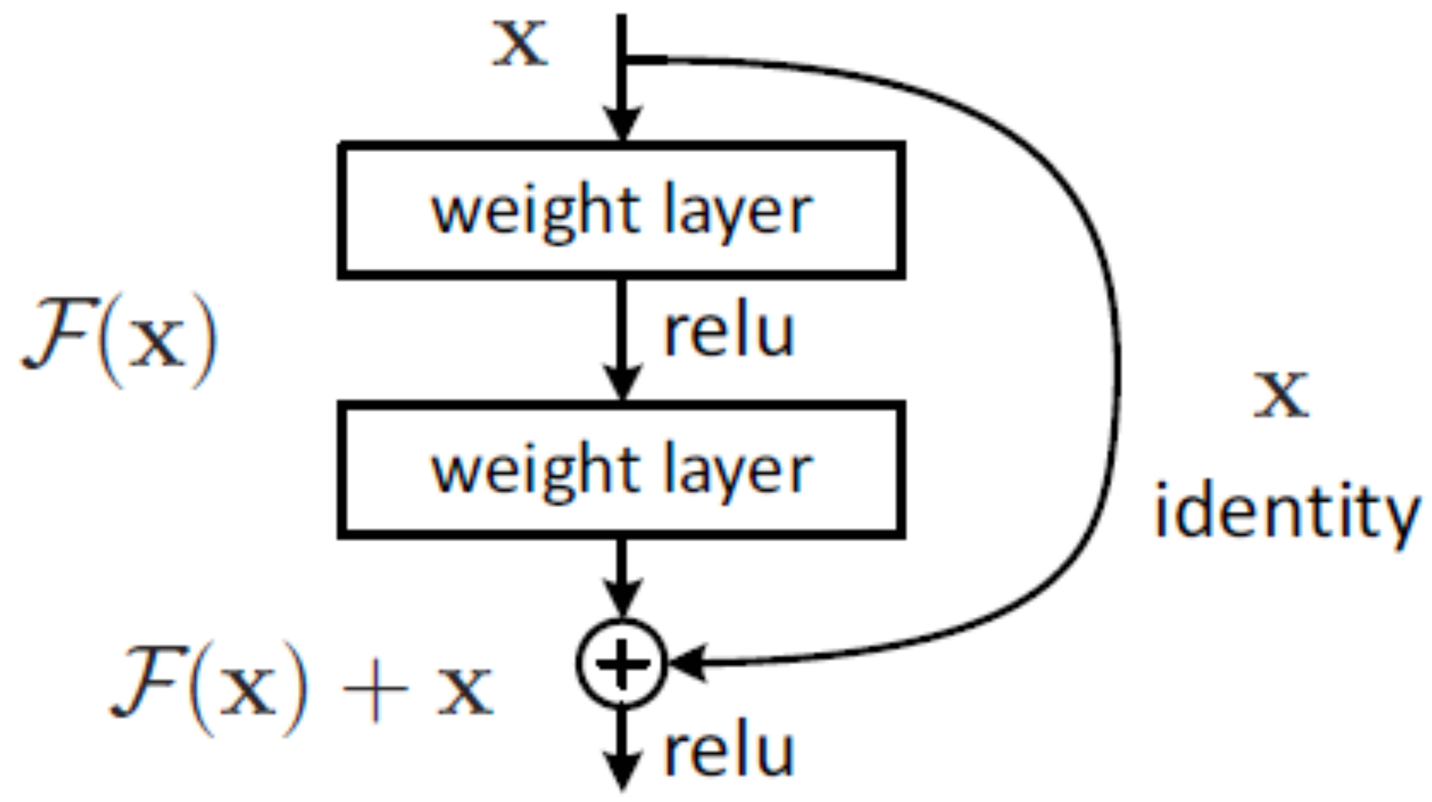


Magnetic Resonance Imaging

How much do we believe our model?

How much do we believe our model?

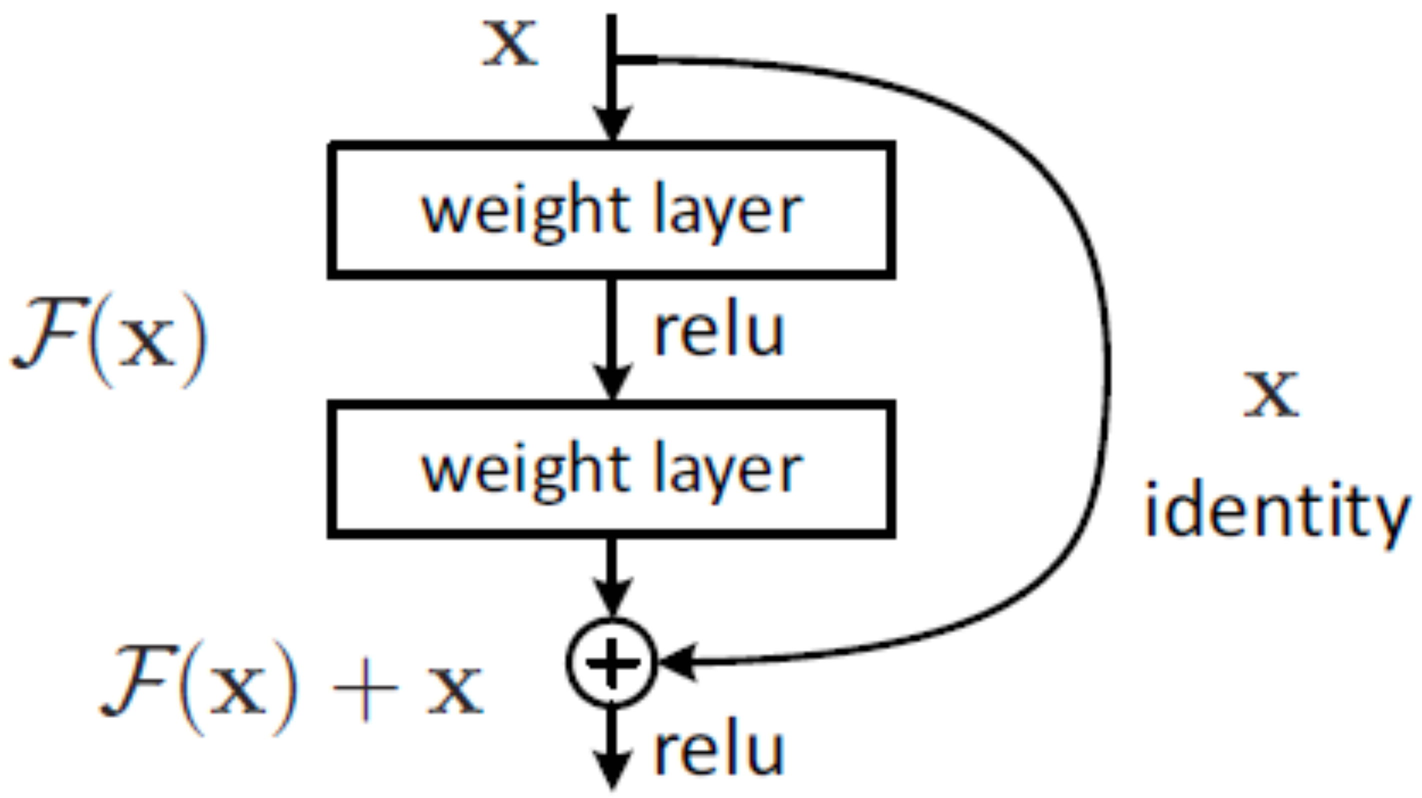
Machine Learning



"What model?"

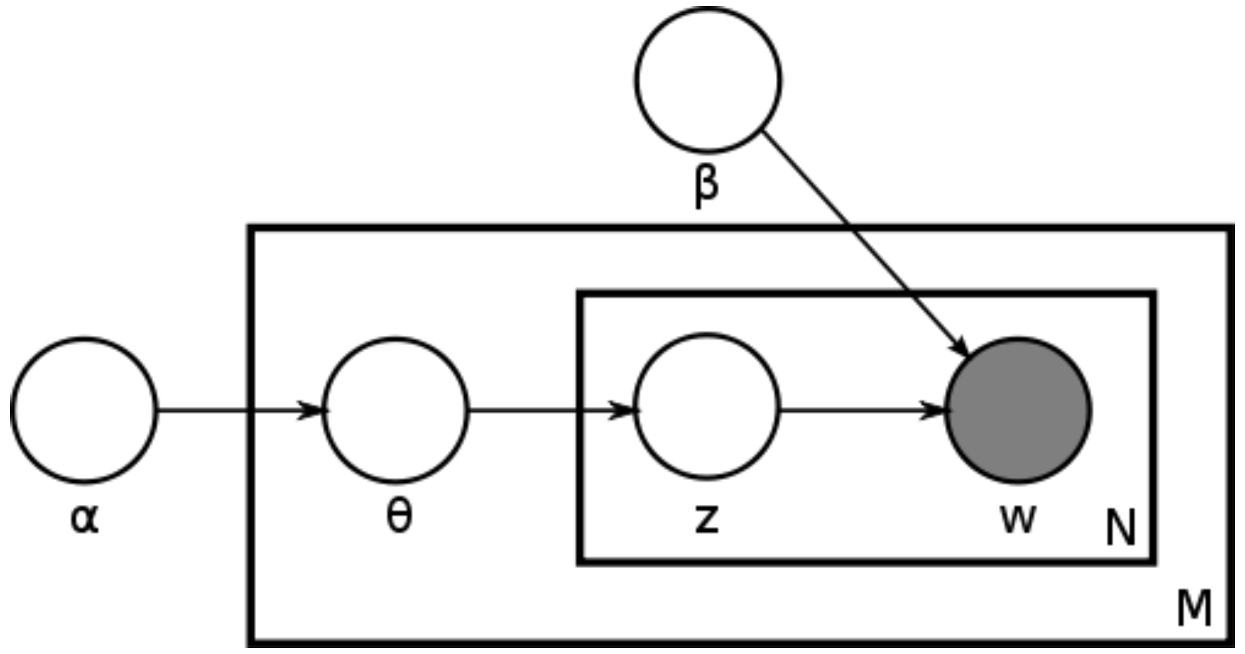
How much do we believe our model?

Machine Learning



"What model?"

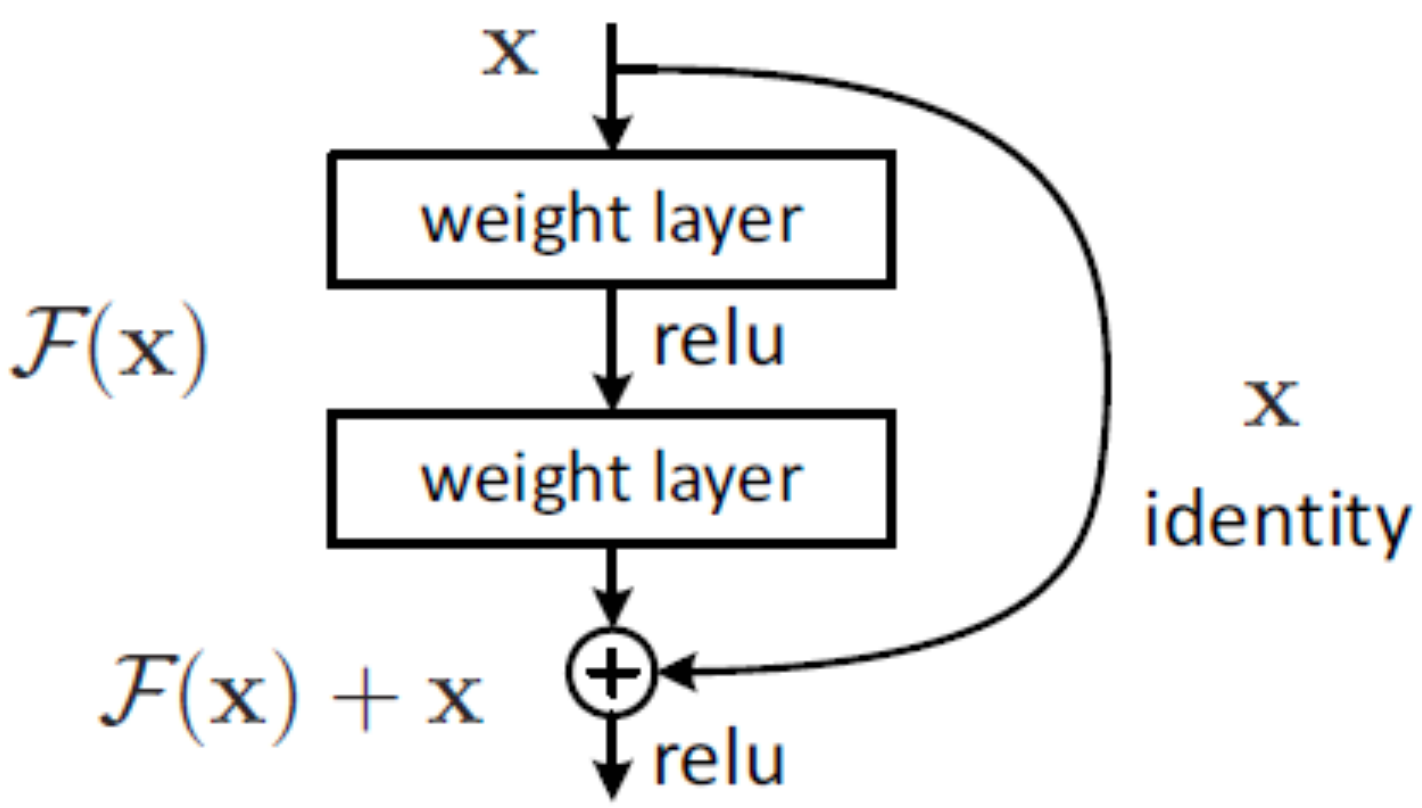
Statistics

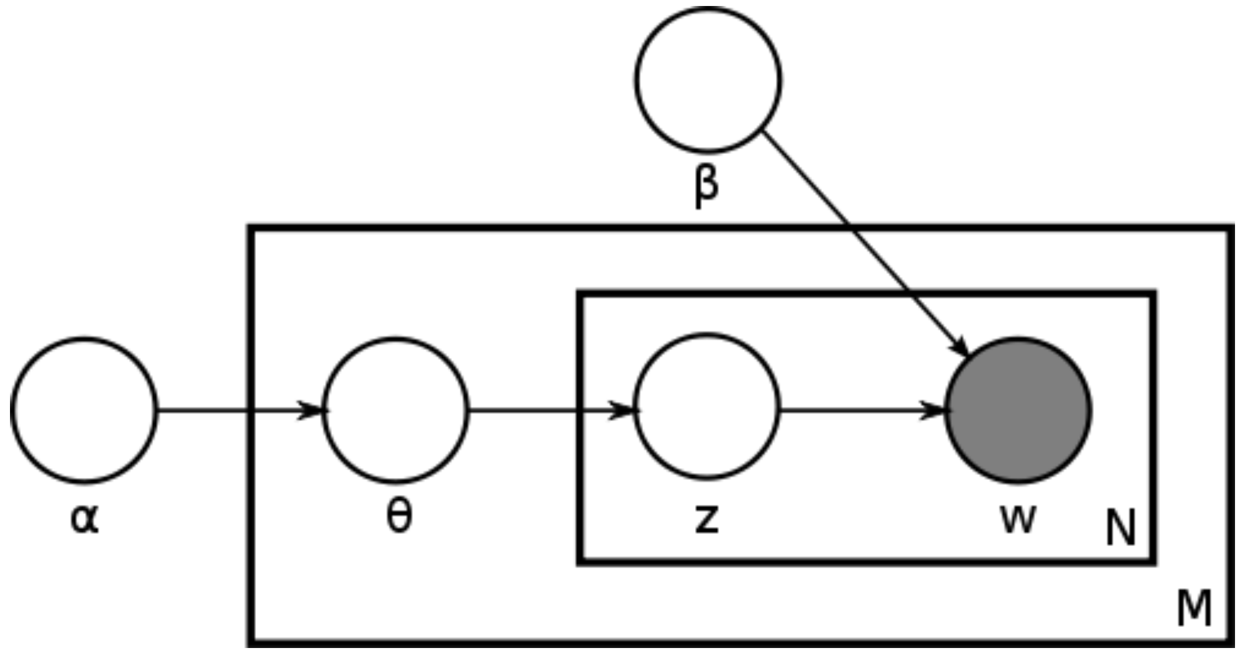


capture some aspect of the system

How much do we believe our model?

Machine Learning	Statistics	Inverse Problems





$$U(P) = \frac{1}{4\pi} \int_S \left[U \frac{\partial}{\partial n} \left(\frac{e^{iks}}{s} \right) - \frac{e^{iks}}{s} \frac{\partial U}{\partial n} \right] dS,$$

(Kirchhoff's diffraction formula)

"What model?"

capture some aspect of the system

Trust completely

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

$$A$$

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

A identity? easy

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

A

identity? easy

not full rank? hard

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

A

identity? easy

not full rank? hard

poor condition number? hard

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

WHY ARE INVERSE PROBLEMS HARD?

Inverse problems are hard for the same reasons that inverting a matrix is hard.

Linear inverse problem

$$y = Ax + \epsilon$$

A

identity? easy

not full rank? hard

poor condition number? hard

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

How to solve? Use prior knowledge!

Google Scholar "deep learning" "inverse problem"

Learning Fast Approximations of Sparse Coding

Karol Gregor and Yann LeCun {KGREGOR,YANN}@CS.NYU.EDU
Courant Institute, New York University, 715 Broadway, New York, NY 10003, USA

Accurate Image Super-Resolution Using Very Deep Convolutional Networks

Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee
Department of ECE, ASRI, Seoul National University, Korea

Image Super-Resolution Using Deep Convolutional Networks

Chao Dong, Chen Change Loy, *Member, IEEE*, Kaiming He, *Member, IEEE*, and Xiaoou Tang, *Fellow, IEEE*

Deep Convolutional Neural Network for Inverse Problems in Imaging

Kyong Hwan Jin, Michael T. McCann, *Member, IEEE*, Emmanuel Froustey, and Michael Unser, *Fellow, IEEE*

ONSAGER-CORRECTED DEEP LEARNING FOR SPARSE LINEAR INVERSE PROBLEMS

Mark Borgerding and Philip Schniter
Dept. of ECE, The Ohio State University, Columbus, OH 43202
Email: borgerding.7@osu.edu, schniter.1@osu.edu

Deeply-Recursive Convolutional Network for Image Super-Resolution

Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee
Department of ECE, ASRI, Seoul National University, Korea

Lensless computational imaging through deep learning

AYAN SINHA[1*], JUSTIN LEE[2], SHUAI LI[1], AND GEORGE BARBASTATHIS[1,3]

LEARNING TO INVERT:
SIGNAL RECOVERY VIA DEEP CONVOLUTIONAL NETWORKS

Ali Mousavi and Richard G. Baraniuk

DEEP CONVOLUTIONAL FRAMELETS: A GENERAL DEEP LEARNING FOR INVERSE PROBLEMS*

JONG CHUL YE*† AND YOSEOB HAN*

One Network to Solve Them All — Solving Linear Inverse Problems using Deep Projection Models

J. H. Rick Chang, Chun-Liang Li, Barnabás Póczos, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan
Carnegie Mellon University, Pittsburgh, PA

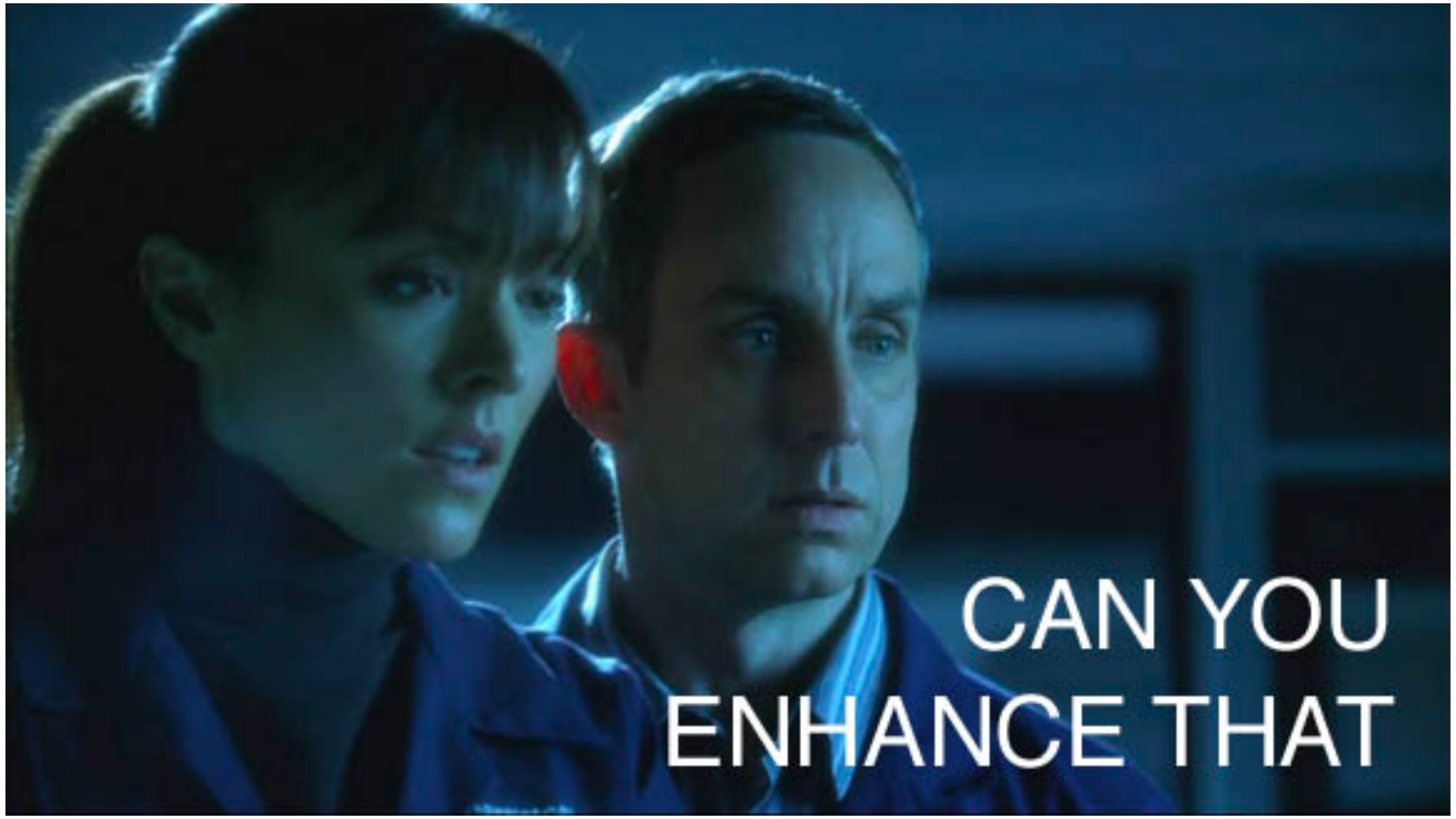
Robust Single Image Super-Resolution via Deep Networks With Sparse Prior

Ding Liu, *Student Member, IEEE*, Zhaowen Wang, *Member, IEEE*, Bihan Wen, *Student Member, IEEE*, Jianchao Yang, *Member, IEEE*, Wei Han, and Thomas S. Huang, *Fellow, IEEE*

AMORTISED MAP INFERENCE FOR IMAGE SUPER-RESOLUTION

Casper Kaae Sønderby[1 2*], Jose Caballero[1], Lucas Theis[1], Wenzhe Shi[1] & Ferenc Huszár[1]
casperkaae@gmail.com, {jcaballero,ltheis,wshi,fhuszar}@twitter.com
[1]Twitter, London, UK
[2]University of Copenhagen, Denmark

#fakenews



Pixel Recursive Super Resolution

Ryan Dahl * Mohammad Norouzi Jonathon Shlens
Google Brain
{rld,mnorouzi,shlens}@google.com

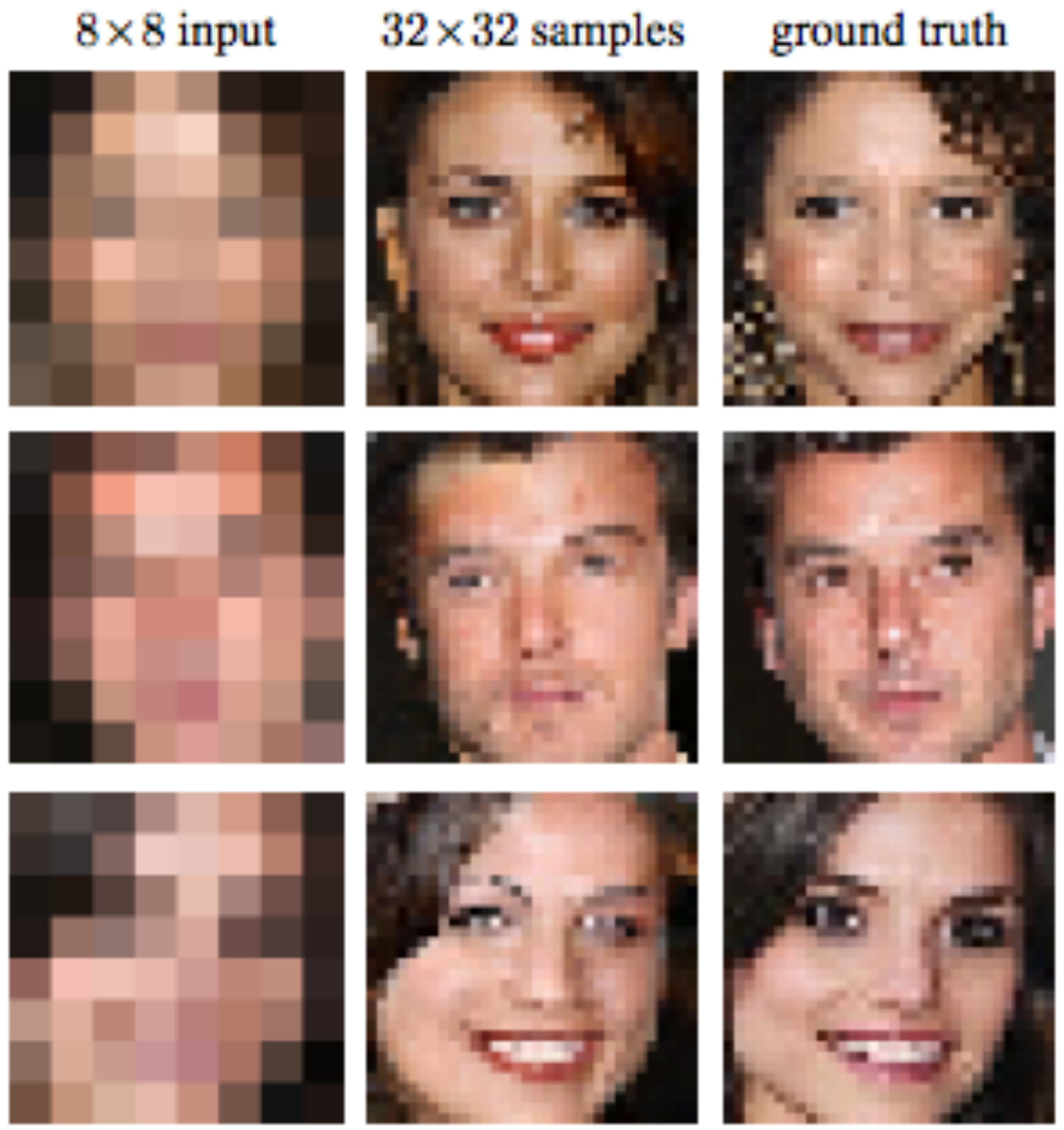


Figure 1: Illustration of our probabilistic pixel recursive super resolution model trained end-to-end on a dataset of celebrity faces. The left column shows 8×8 low resolution inputs from the test set. The middle and last columns show 32×32 images as predicted by our model *vs.* the ground truth. Our model incorporates strong face priors to synthesize realistic hair and skin details.

Is this invertable?

Is this invertable?

CAN ONE HEAR THE SHAPE OF A DRUM?

MARK KAC, The Rockefeller University, New York

To George Eugene Uhlenbeck on the occasion of his sixty-fifth birthday

> “La Physique ne nous donne pas seulement l’occasion de résoudre des problèmes . . . , elle nous fait presentir la solution.” H. POINCARÉ.

Before I explain the title and introduce the theme of the lecture I should like to state that my presentation will be more in the nature of a leisurely excursion than of an organized tour. It will not be my purpose to reach a specified destination at a scheduled time. Rather I should like to allow myself on many occasions the luxury of stopping and looking around. So much effort is being spent on streamlining mathematics and in rendering it more efficient, that a solitary transgression against the trend could perhaps be forgiven.

American Mathematical Monthly 1966

Is this invertable?

CAN ONE HEAR THE SHAPE OF A DRUM?

MARK KAC, The Rockefeller University, New York

To George Eugene Uhlenbeck on the occasion of his sixty-fifth birthday

> "La Physique ne nous donne pas seulement l'occasion de résoudre des problèmes . . . , elle nous fait presentir la solution." H. POINCARÉ.

Before I explain the title and introduce the theme of the lecture I should like to state that my presentation will be more in the nature of a leisurely excursion than of an organized tour. It will not be my purpose to reach a specified destination at a scheduled time. Rather I should like to allow myself on many occasions the luxury of stopping and looking around. So much effort is being spent on streamlining mathematics and in rendering it more efficient, that a solitary transgression against the trend could perhaps be forgiven.

American Mathematical Monthly 1966

In 1966, in a celebrated paper (Kac, 1966), Mark Kac formulated the famous question "Can one hear the shape of a drum?". This provocative question is of course to be understood mathematically as follows: Is it possible to find two (or more) non-isometric Euclidean simply connected domains for which the sets $\{E_n \parallel n \in \mathbb{N}\}$ of solutions of (1) with $\Psi_{|\text{Boundary}} = 0$ are identical? More broadly, the question raises the issue of the inverse problem of retrieving information about a drum from knowledge of its spectral properties. As the spectroscopist A. Schuster put it in an 1882 report to the British Association for the Advancement of Science: "To find out the different tunes sent out by a vibrating system is a problem which may or may not be solvable in certain special cases, but it would baffle the most skillful mathematicians to solve the inverse problem and to find out the shape of a bell by means of the sounds which it is capable of sending out. And this is the problem which ultimately spectroscopy hopes to solve in the case of light. In the meantime we must welcome with delight even the smallest step in the desired direction." (Mehra and Rechenberg, 2000). Actually, it was known very early, from Weyl's formula, that one can "hear" the area of a drum and the length of its perimeter (see section V.A, and (Vaa *et al.*, 2005) for a historical account of the problem). But could the shape itself be retrieved from the spectrum? That is, what kind of information on the geometry is it possible to gather from the knowledge of the spectrum, for instance, using semiclassical methods that allow investigation of the quantum-classical correspondence? And what kind of sufficient conditions allow the geometry to be entirely specified from the spectrum?

Is this invertable?

In 1966, in a celebrated paper (Kac, 1966), Mark Kac formulated the famous question "Can one hear the shape of a drum?". This provocative question is of course to be understood mathematically as follows: Is it possible to find two (or more) non-isometric Euclidean simply connected domains for which the sets $\{E_n \| n \in \mathbb{N}\}$ of solutions of

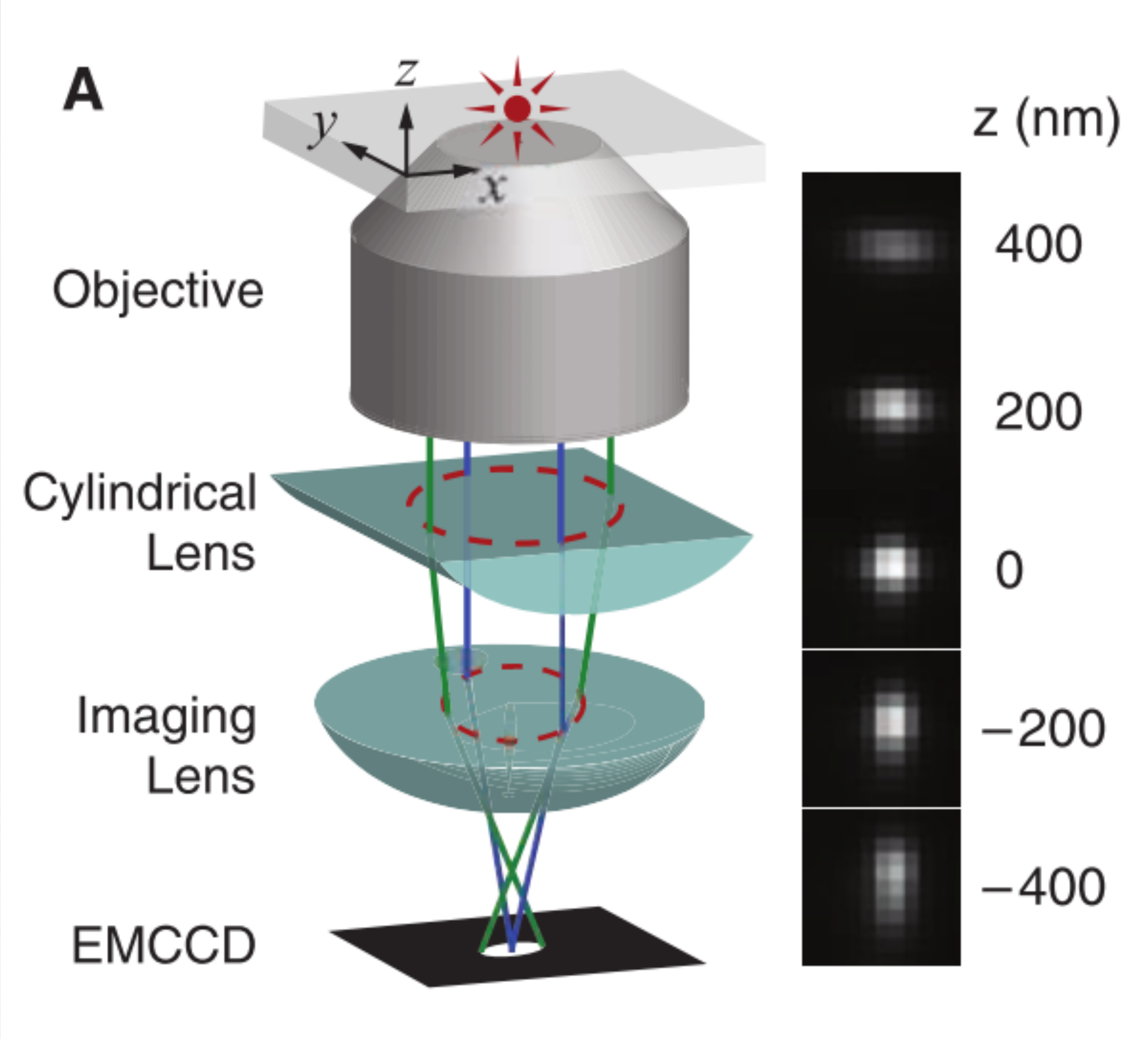
its spectral properties. As the spectroscopist A. Schuster put it in an 1882 report to the British Association for the Advancement of Science: "To find out the different tunes sent out by a vibrating system is a problem which may or may not be solvable in certain special cases, but it would baffle the most skillful mathematicians to solve the inverse problem and to find out the shape of a bell by means of the sounds which it is capable of sending out. And this is the problem which ultimately spectroscopy hopes to solve in the case of light. In the meantime we must welcome with delight even the smallest step in the desired direction." (Mehra and Rechenberg, 2000).

the quantum-classical correspondence? And what kind of sufficient conditions allow the geometry to be entirely specified from the spectrum?

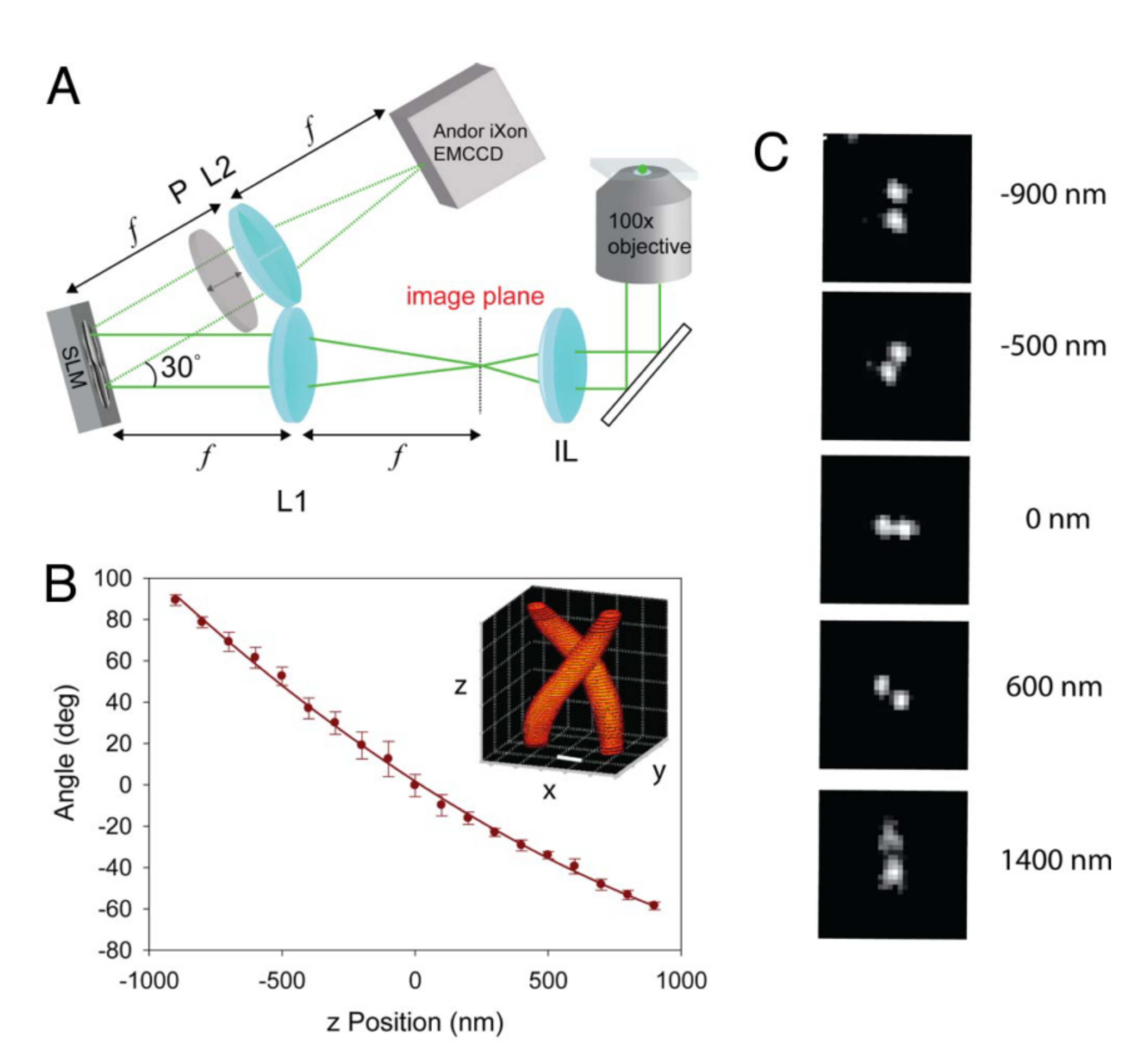
American

ENGINEERED PSF FOR 3D STORM

Astigmatic



Double Helix



Huang, B., Wang, W., Bates, M., & Zhuang, X. (2008). Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy. Science, 319(5864), 810–813. http://doi.org/10.1126/science.1153529

Pavani, S. R. P., Thompson, M. A., Biteen, J. S., Lord, S. J., Liu, N., Twieg, R. J., … Moerner, W. E. (2009). Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. Proceedings of the National Academy of Sciences of the United States of America, 106(9), 2995–2999. http://doi.org/10.1073/pnas.0900245106

LEARNING FROM SIMULATION

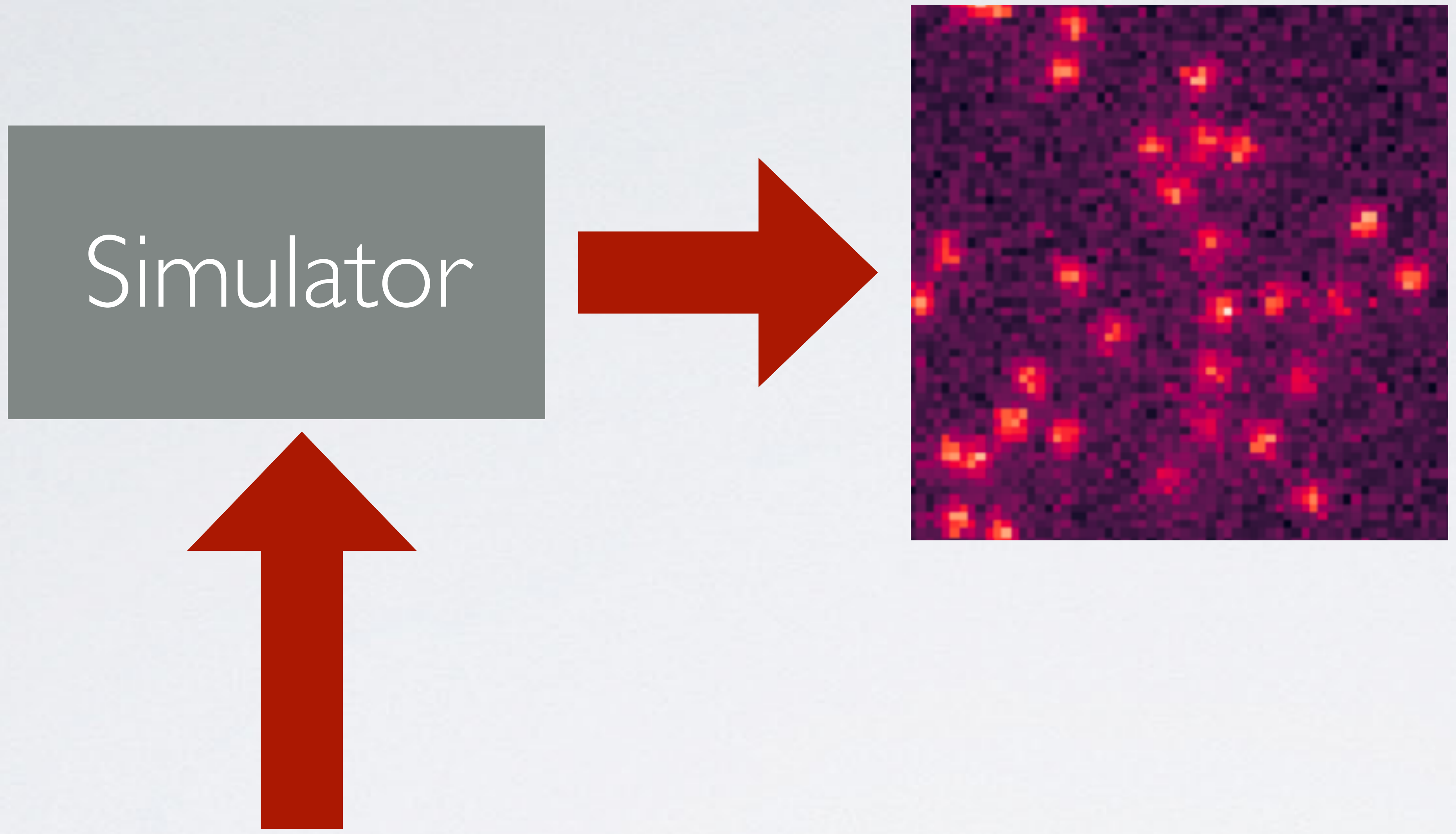
LEARNING FROM SIMULATION

$$\gamma = \{(w_j, \theta_j), \cdots\}$$

LEARNING FROM SIMULATION

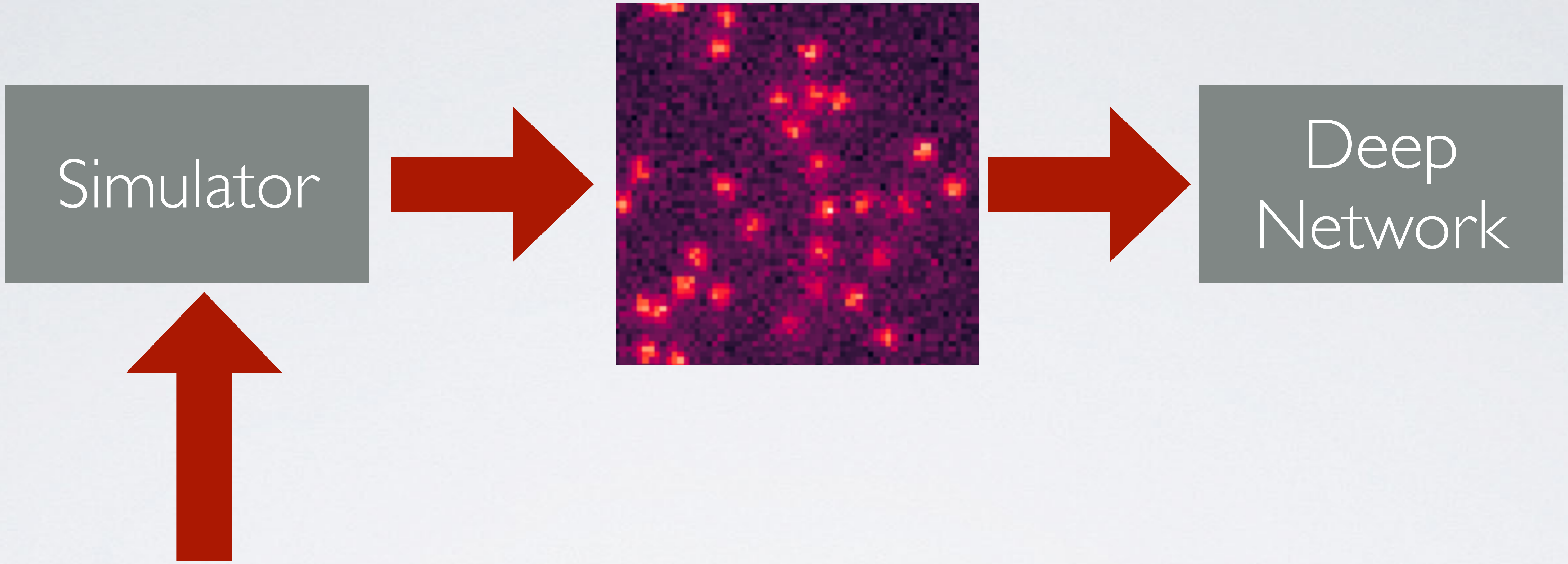
$$\gamma = \{(w_j, \theta_j), \cdots\}$$

LEARNING FROM SIMULATION



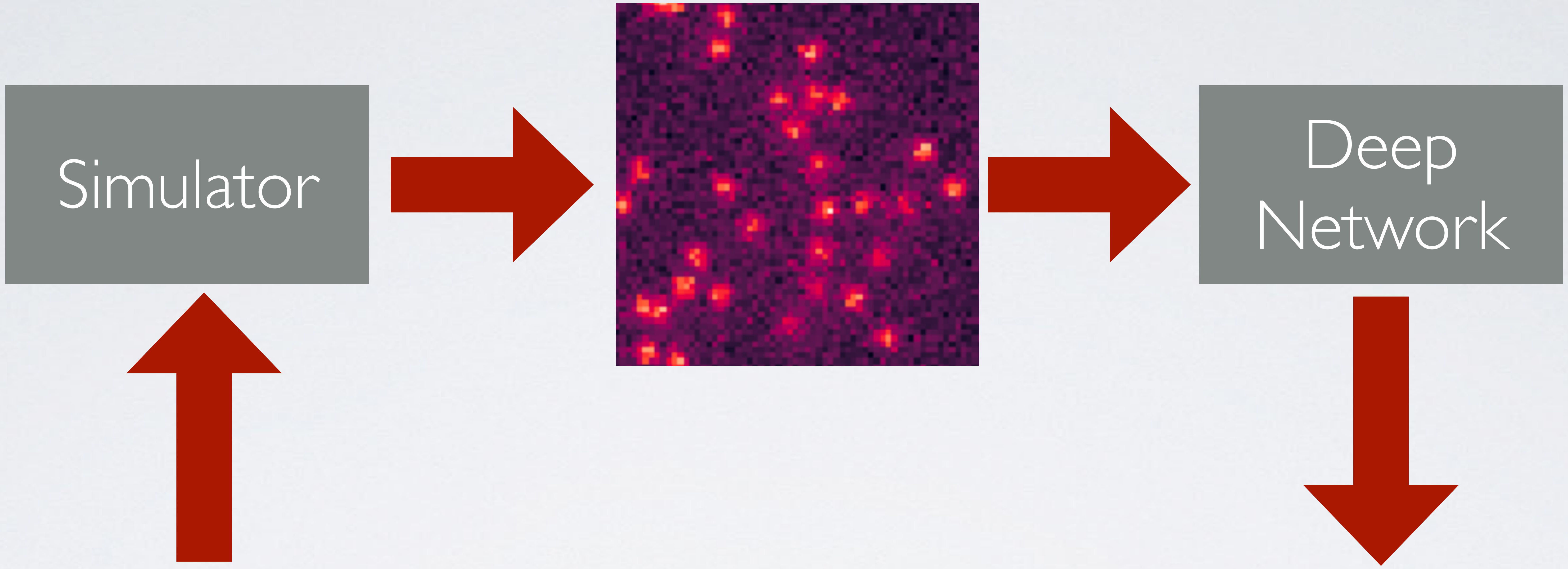
$$\gamma = \{(w_j, \theta_j), \cdots\}$$

LEARNING FROM SIMULATION



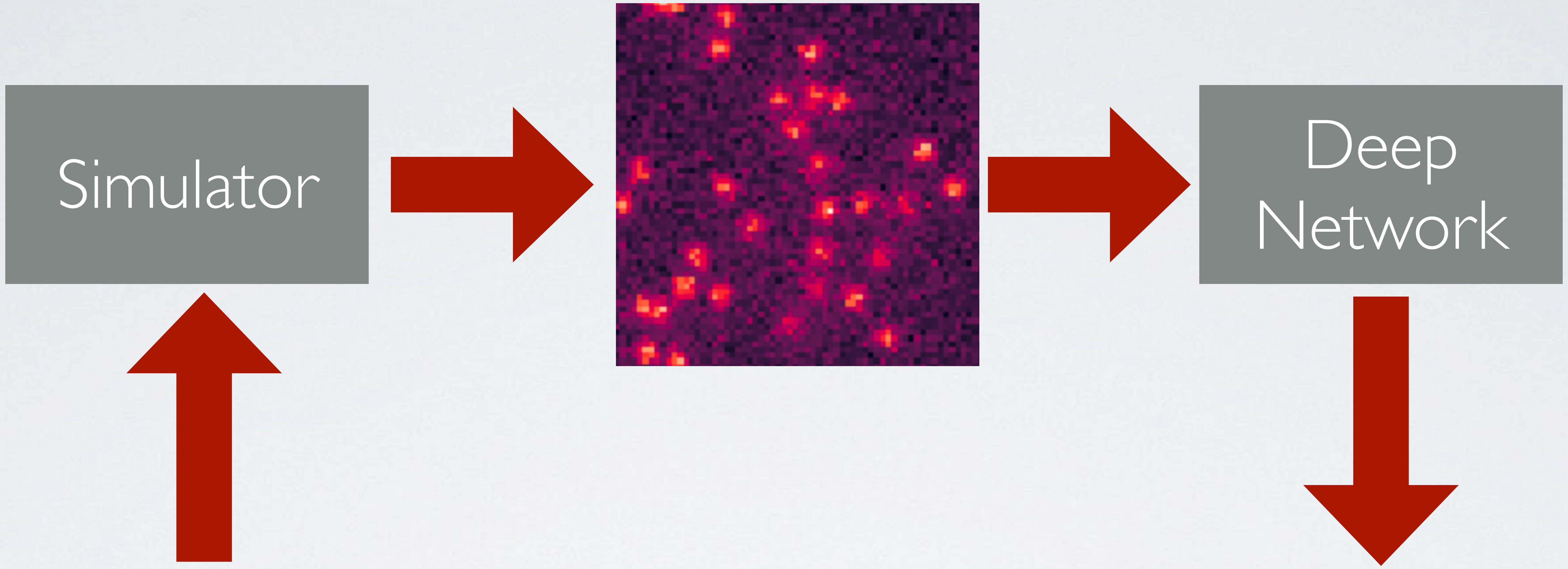
$$\gamma = \{(w_j, \theta_j), \cdots\}$$

LEARNING FROM SIMULATION



$$\gamma = \{(w_j, \theta_j), \cdots\}$$

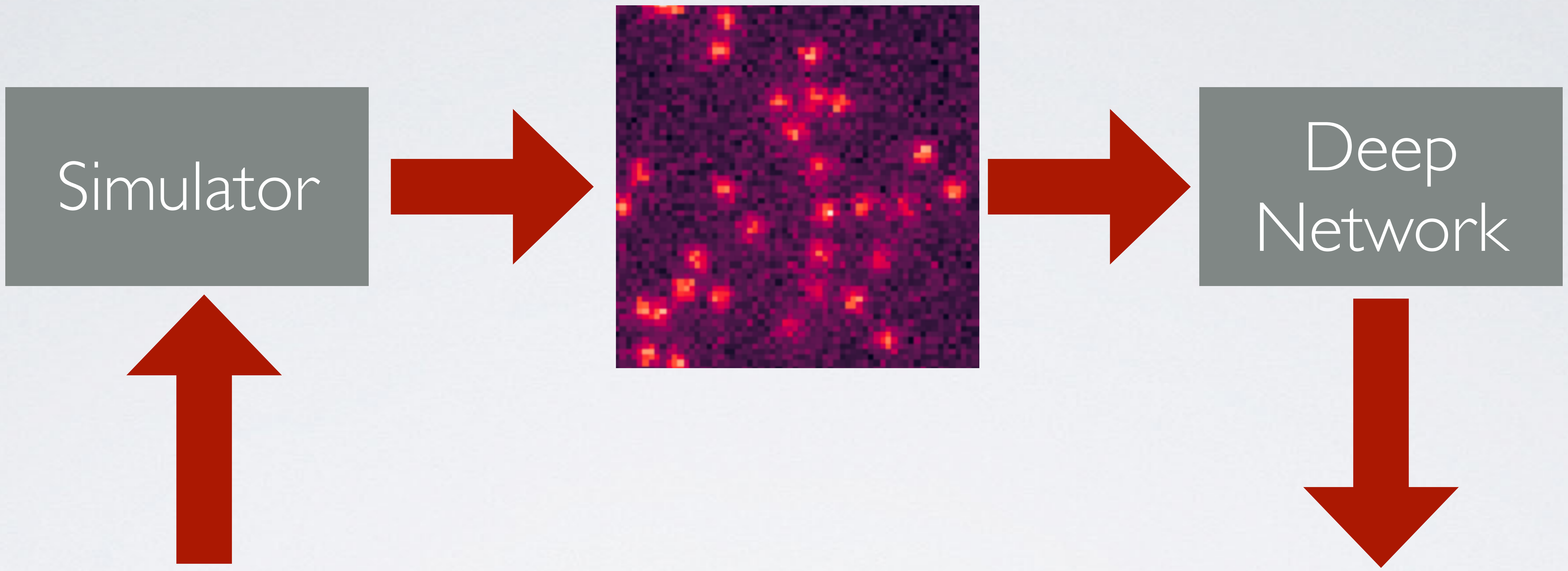
LEARNING FROM SIMULATION



$\gamma = \{(w_j, \theta_j), \cdots\}$

$\hat{\gamma} = \{(w_i, \theta_i), \cdots\}$

LEARNING FROM SIMULATION

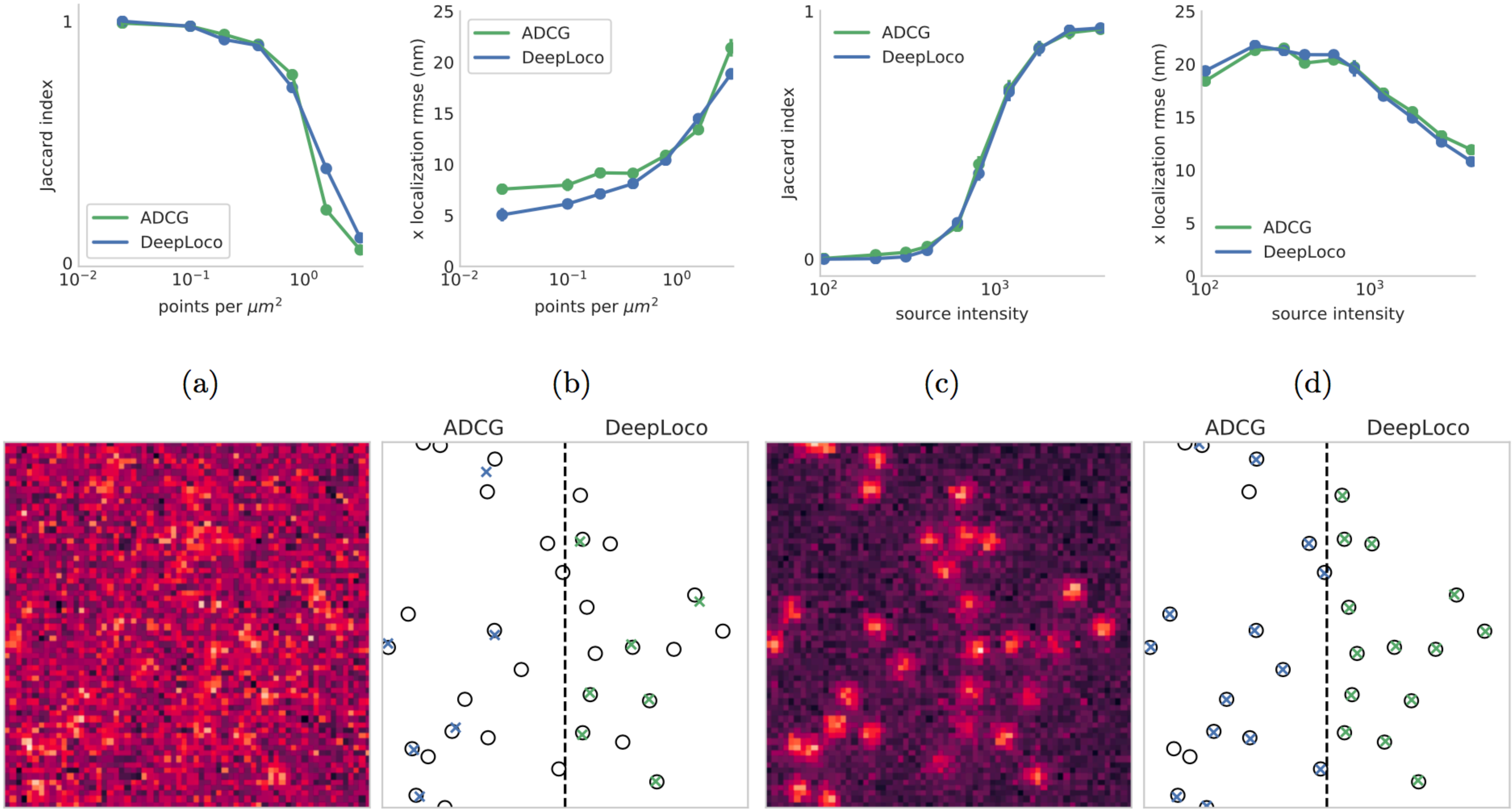


$\gamma = \{(w_j, \theta_j), \cdots\}$

$\hat{\gamma} = \{(w_i, \theta_i), \cdots\}$

$$\ell(\hat{\gamma}, \gamma) = ||\hat{I}_{\hat{\gamma}}(x) - I_{\gamma}(x)||_2^2$$

2D COMPARISON



ADCG: Alternating Descent Conditional Gradient

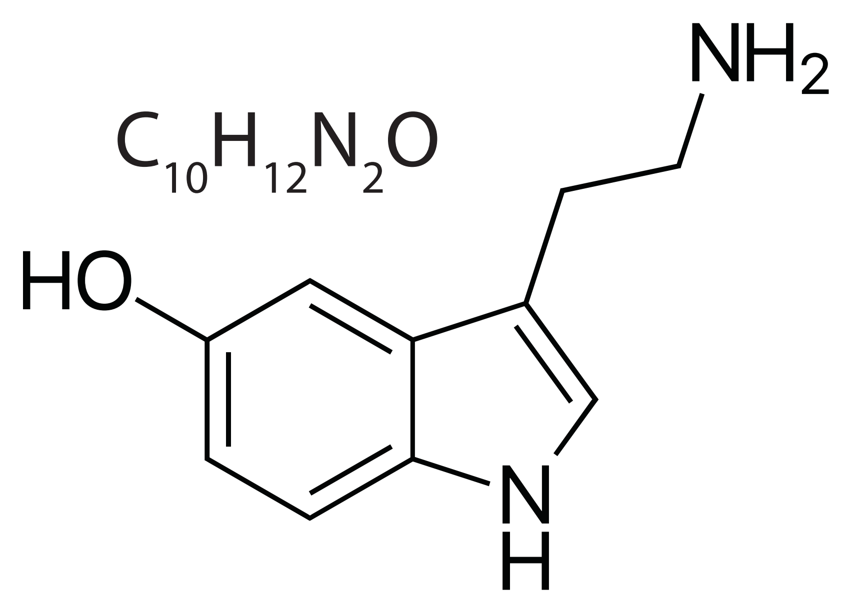
Boyd, Nicholas, Geoffrey Schiebinger, and Benjamin Recht. "The alternating descent conditional gradient method for sparse inverse problems." *SIAM Journal on Optimization* 27, no. 2 (2017): 616-639.

Beyond Images to Graphs

Solving Spectroscopic Inverse Problems

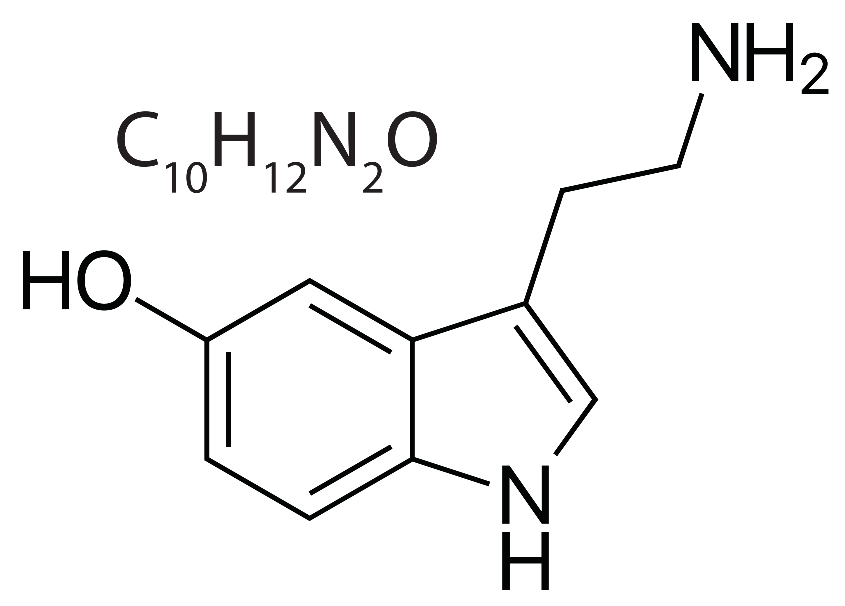
Beyond Images to Graphs

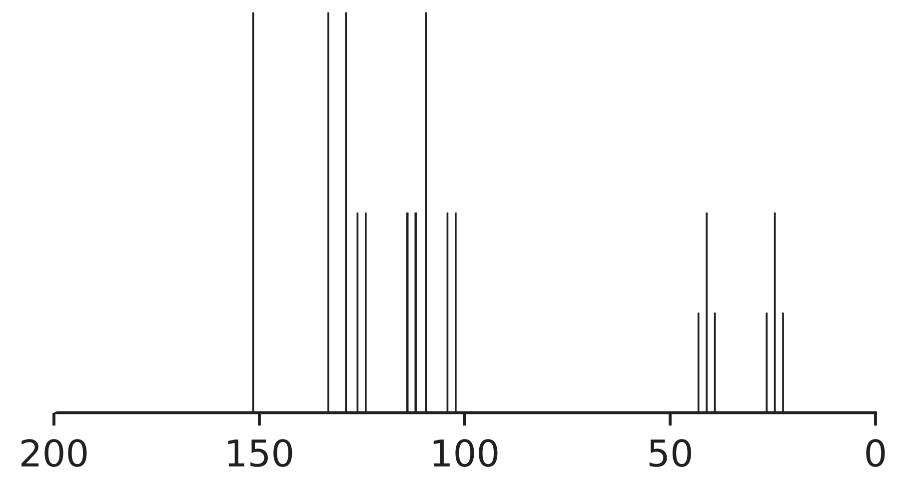
Solving Spectroscopic Inverse Problems



Beyond Images to Graphs

Solving Spectroscopic Inverse Problems



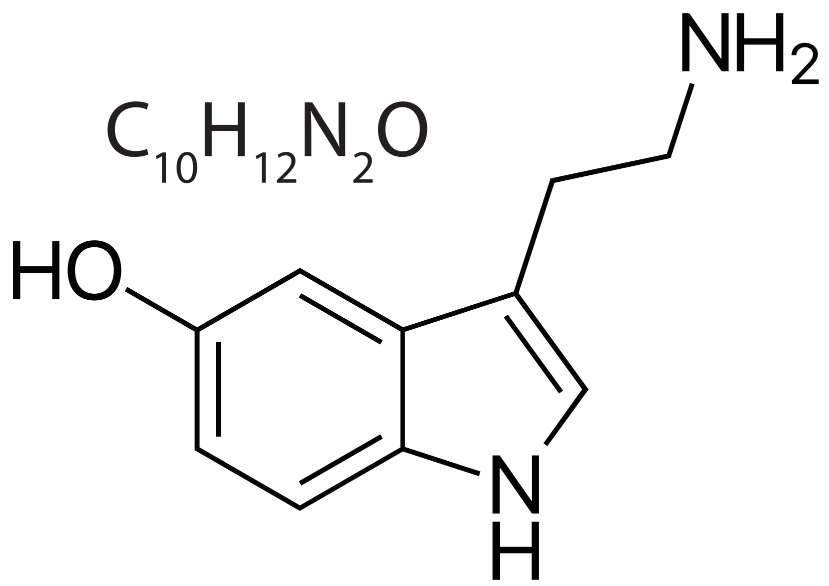


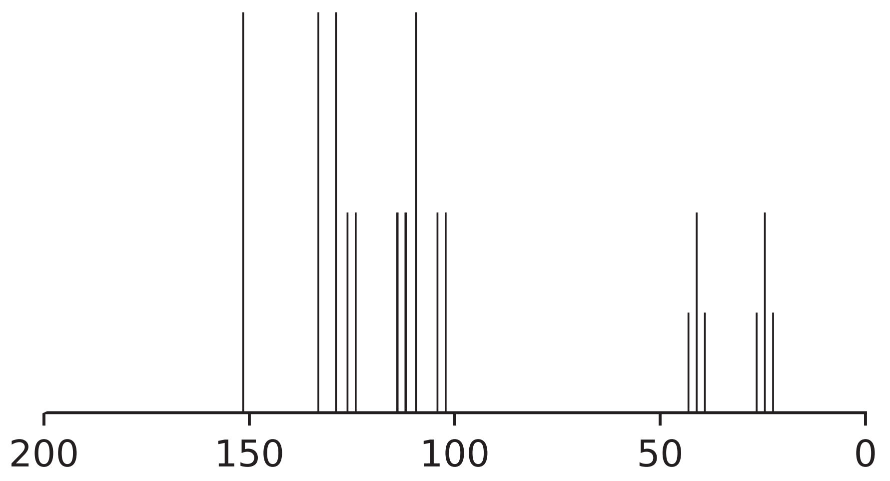
Beyond Images to Graphs

Solving Spectroscopic Inverse Problems

The **forward** problem

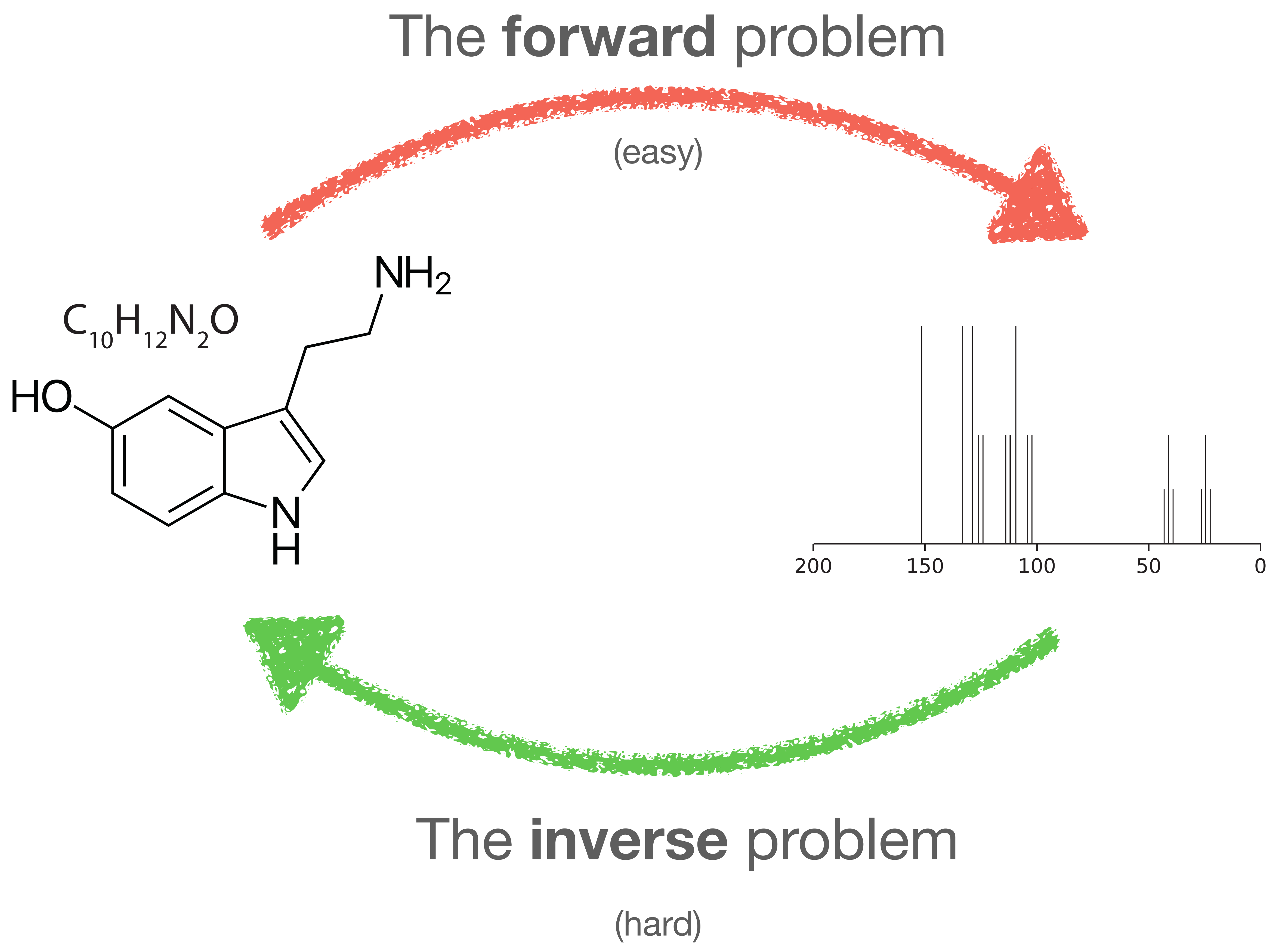
(easy)





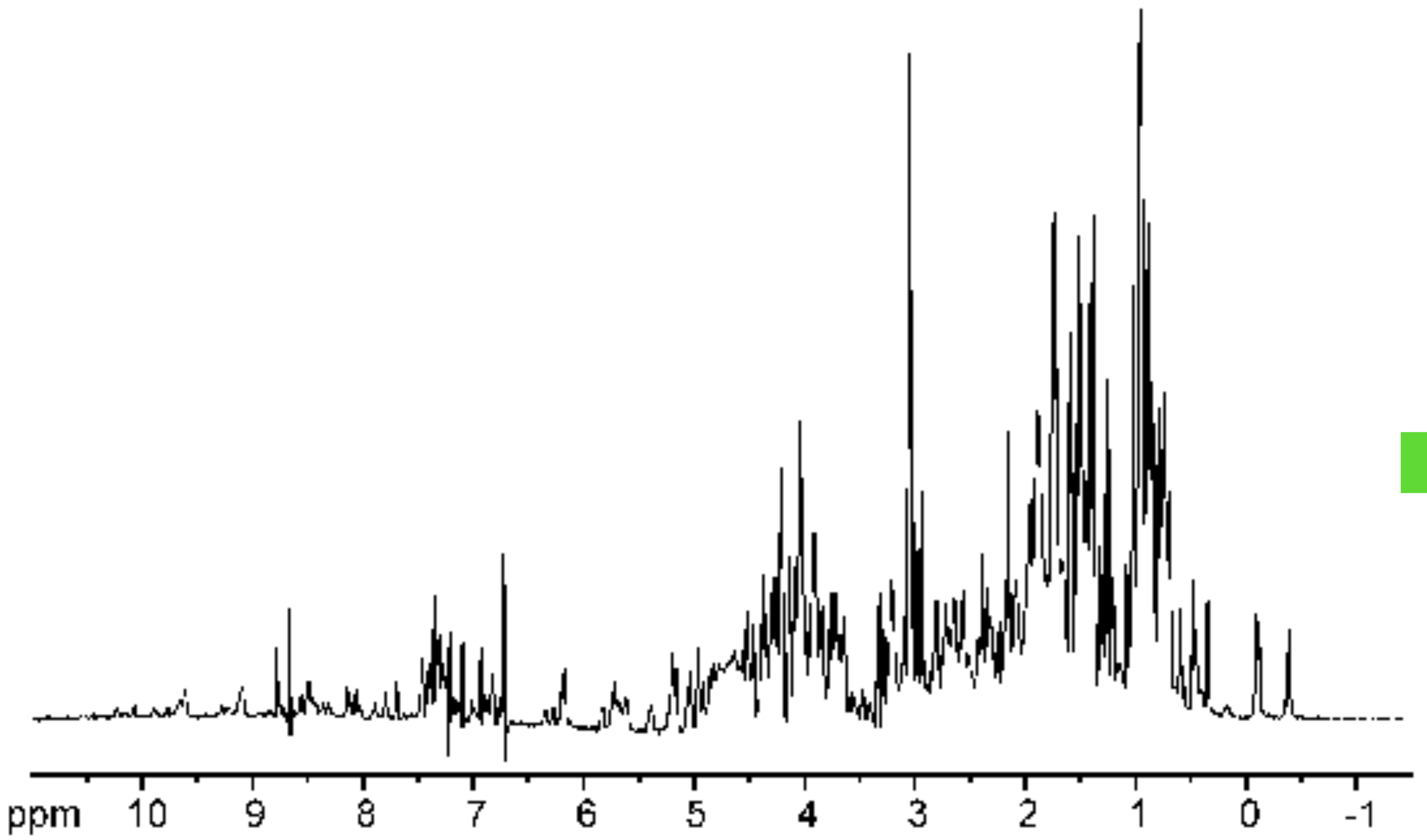
Beyond Images to Graphs

Solving Spectroscopic Inverse Problems

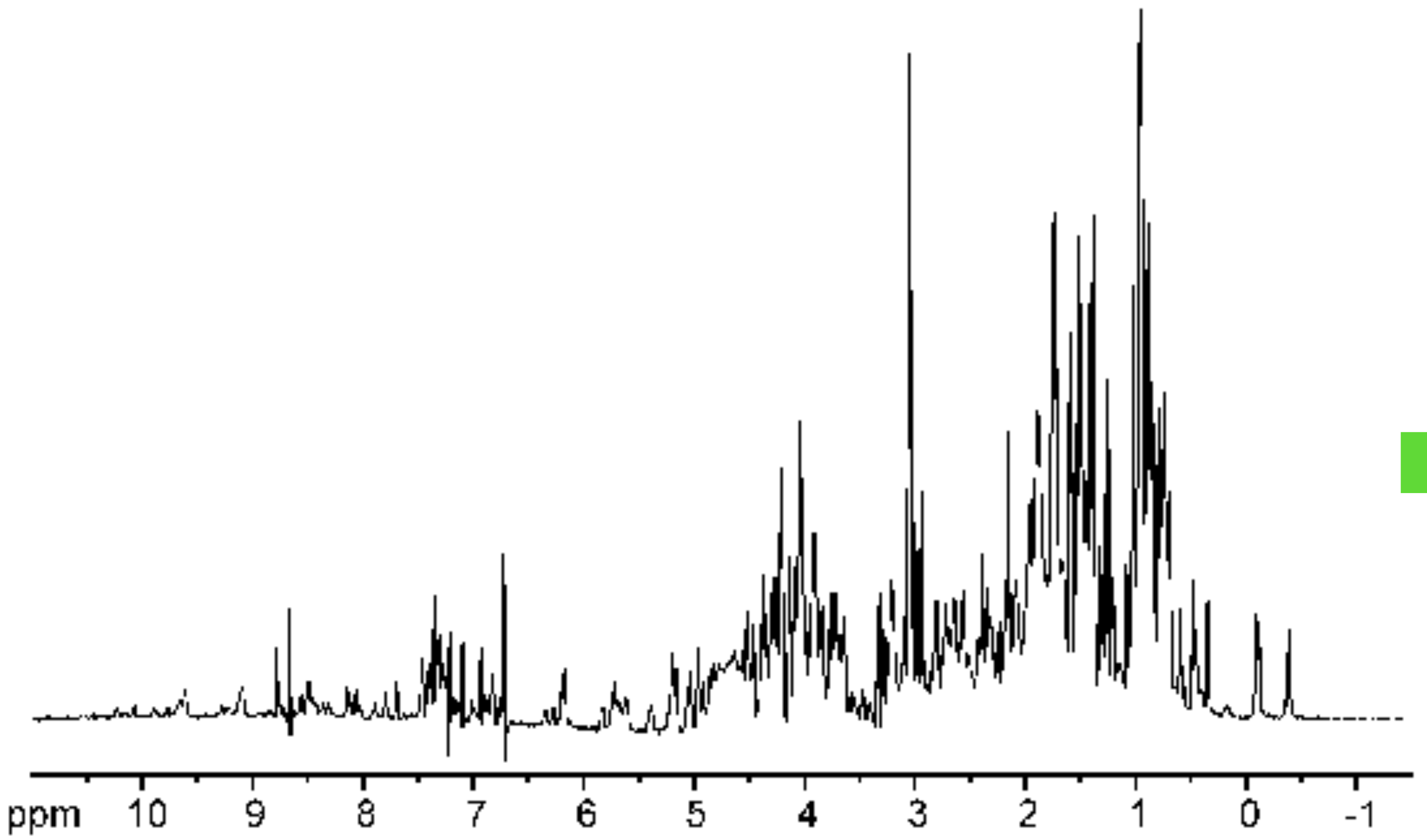


Beyond Images to Graphs

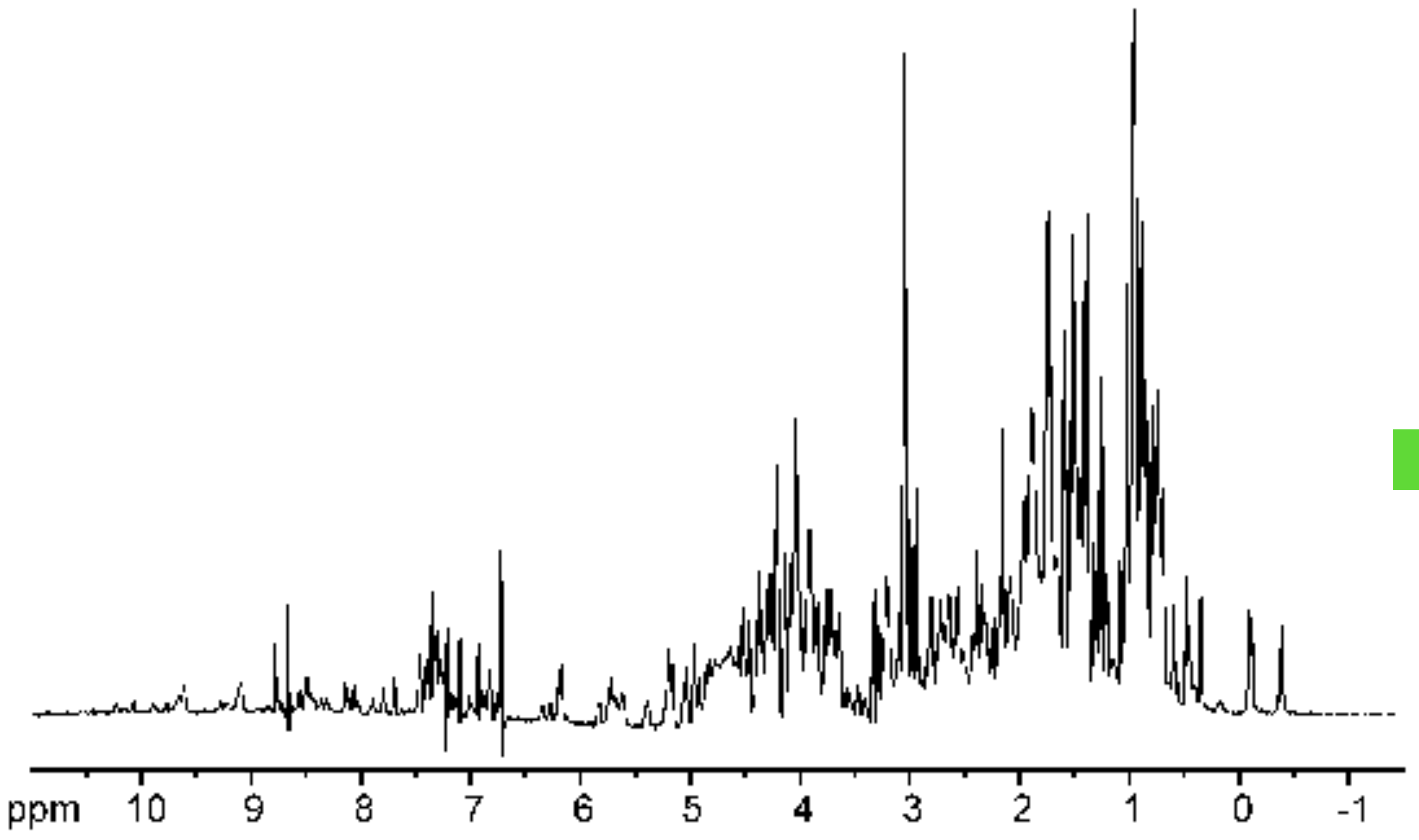
Beyond Images to Graphs

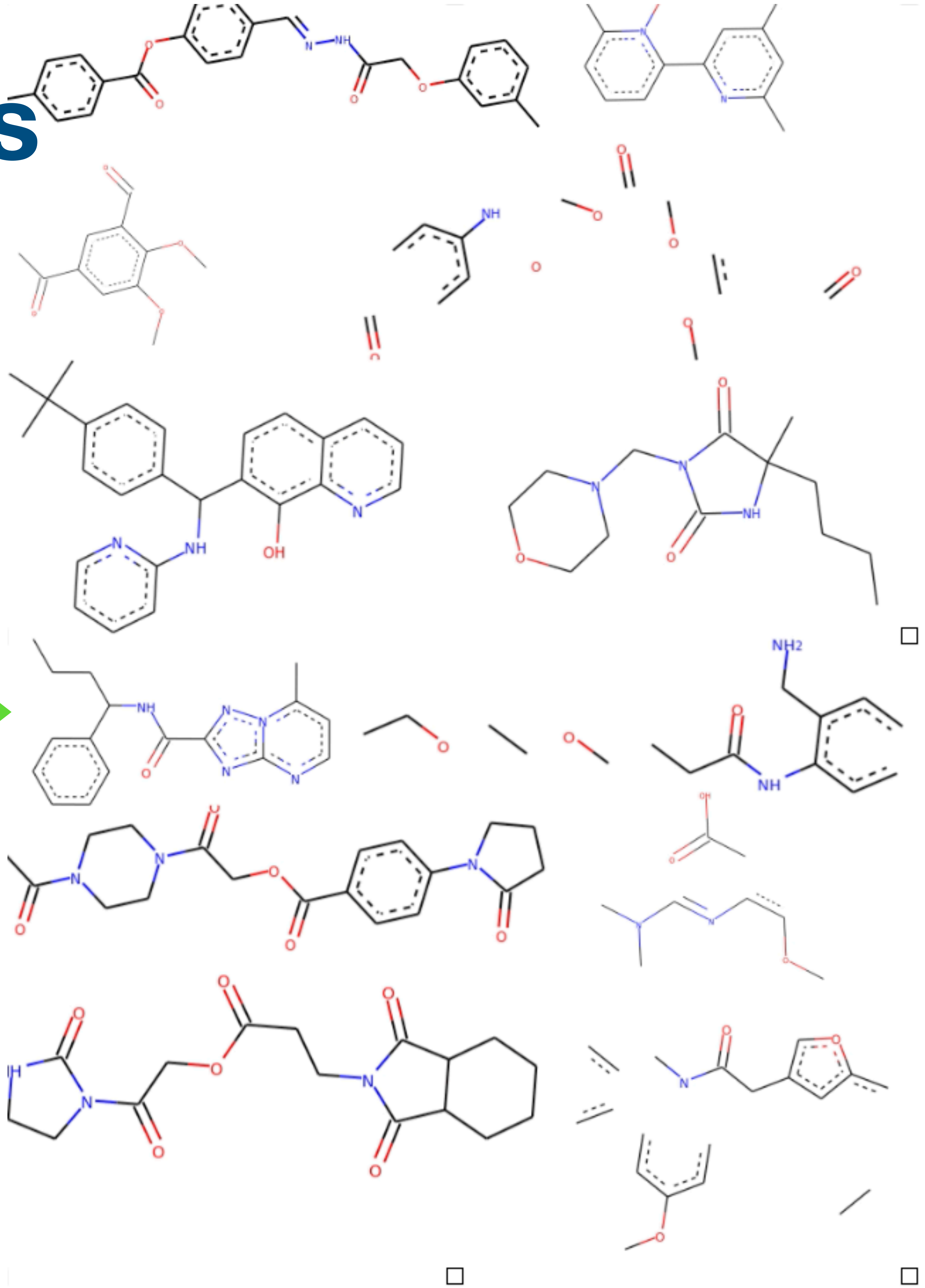


Beyond Images to Graphs



Beyond Images to Graphs





Computational Measurement

How can we design new measurement systems
to be more interpretable / useful for AI?

What is your query?

How many bits are you trying to extract from your system?

What is your query?

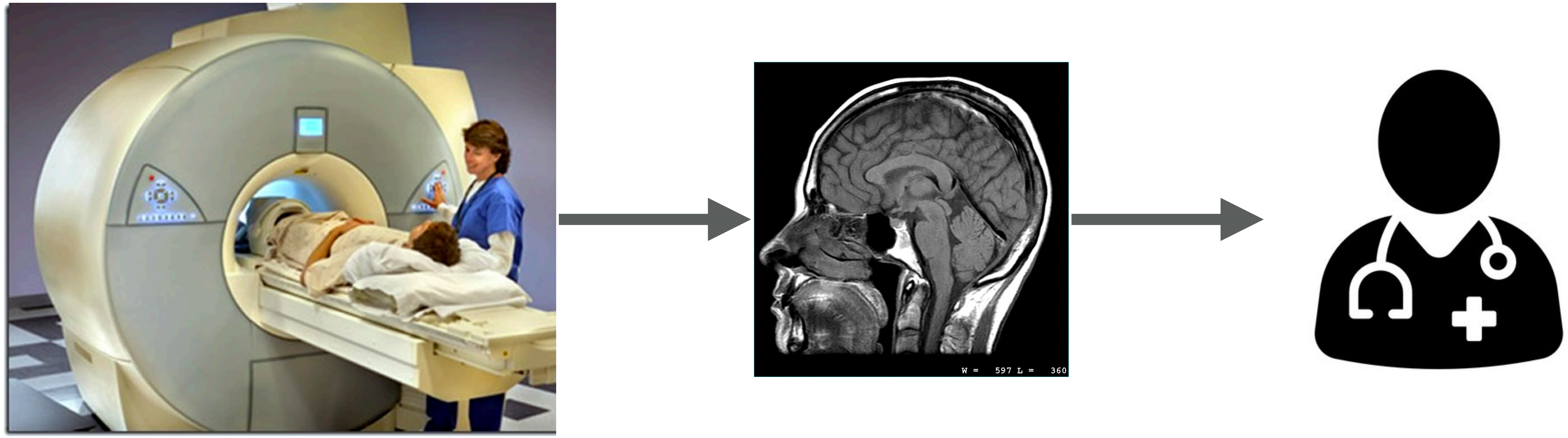
How many bits are you trying to extract from your system?

What is your query?

How many bits are you trying to extract from your system?

What is your query?

How many bits are you trying to extract from your system?



What is your query?

How many bits are you trying to extract from your system?

What is your query?

How many bits are you trying to extract from your system?

1-bit measurement system

Tumor?

[ ] Yes
[ ] No

What is your query?

How many bits are you trying to extract from your system?

1-bit measurement system

Tumor?

[ ] Yes
[ ] No

12 Megapixels (MB)

What is your query?

How many bits are you trying to extract from your system?

1-bit measurement system

Tumor?

[ ] Yes

[ ] No

12 Megapixels (MB)

JPEG

170 kB

What is your query?

How many bits are you trying to extract from your system?

1-bit measurement system

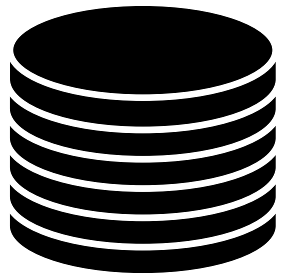
Tumor?

[ ] Yes

[ ] No

12 Megapixels (MB)

JPEG

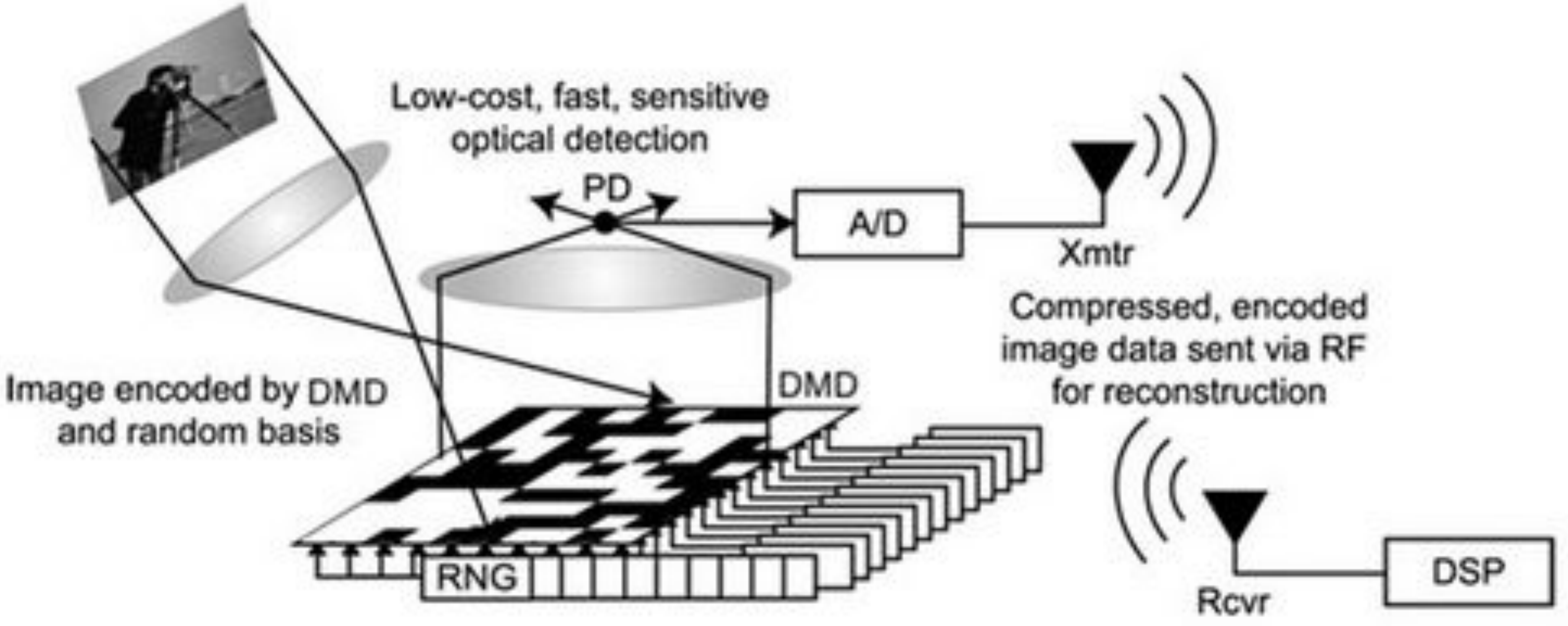


170 kB

Is it possible to only collect 170 kB of data?

Compressive Sensing

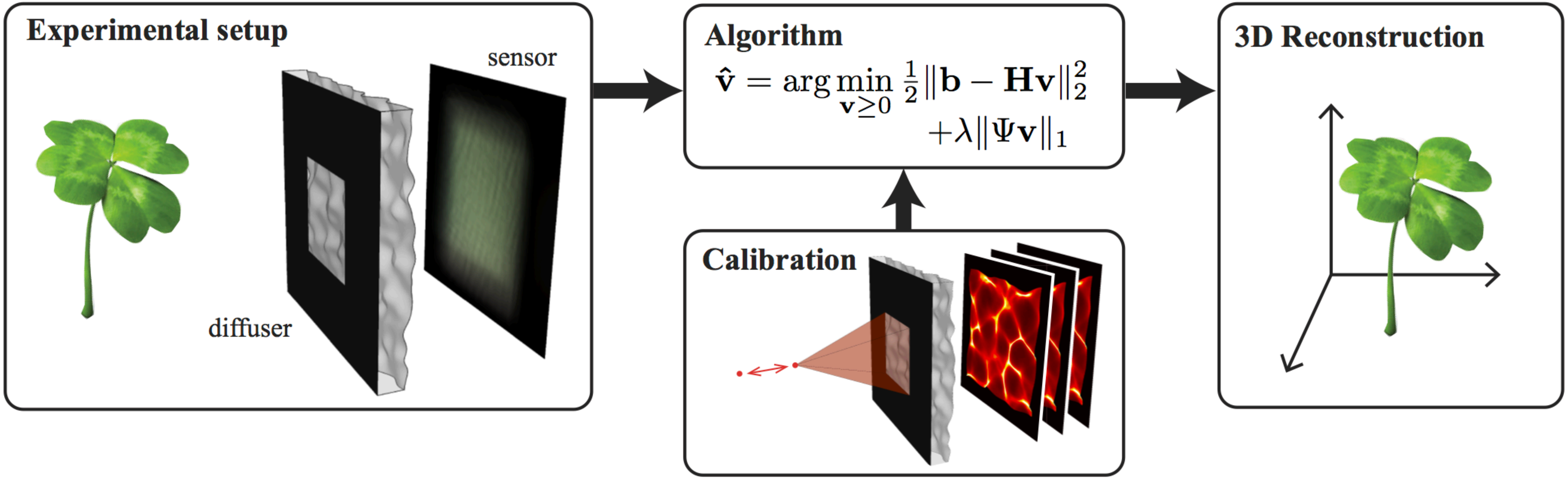
IEEE Signal Processing Magazine March 2008



[FIG2] Single-pixel photo album. (a) 256×256 conventional image of a black-and-white R. (b) Single-pixel camera reconstructed image from $M = 1,300$ random measurements ($50\times$ sub-Nyquist). (c) 256×256 pixel color reconstruction of a printout of the Mandrill test image imaged in a low-light setting using a single photomultiplier tube sensor, RGB color filters, and $M = 6,500$ random measurements.

DiffuserCam

Single-shot 3D acquisition



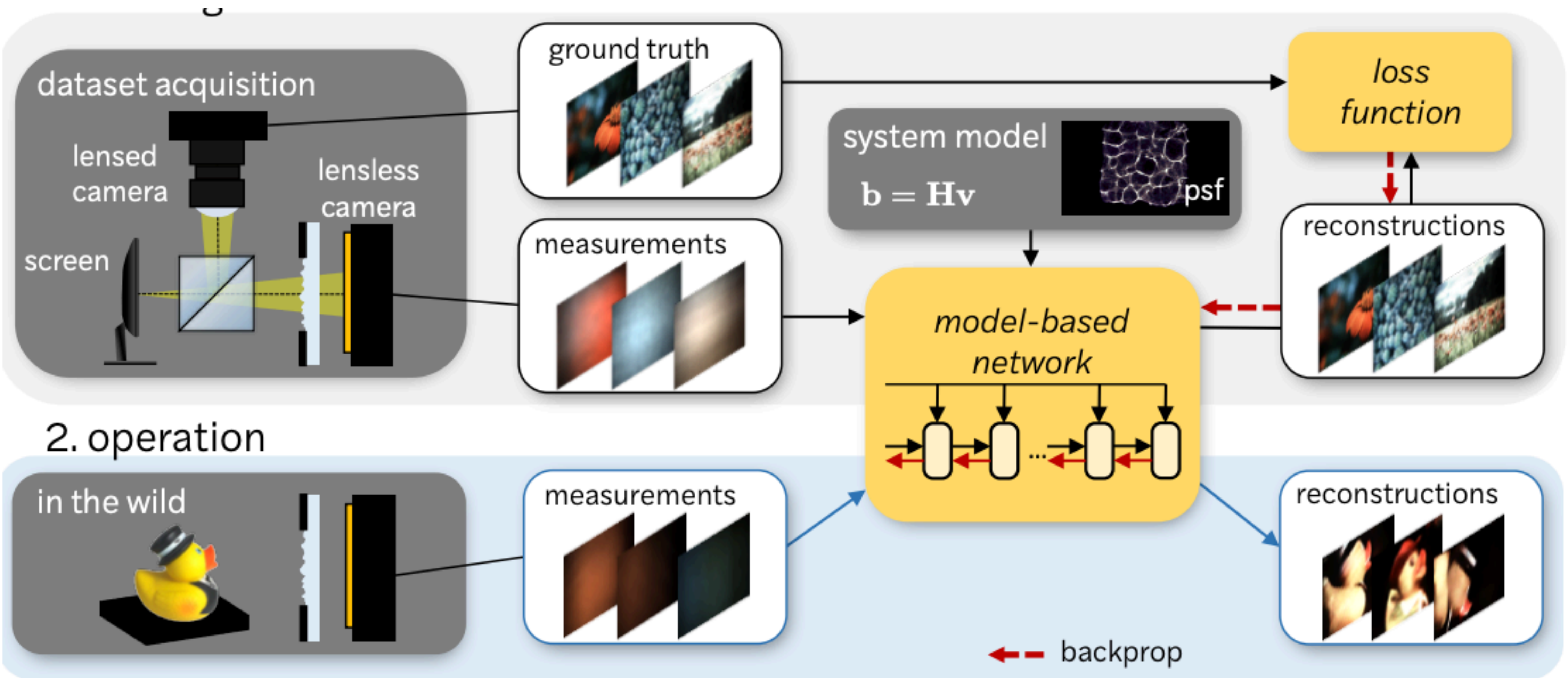
Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica 5, 1-9 (2018)

DiffuserCam

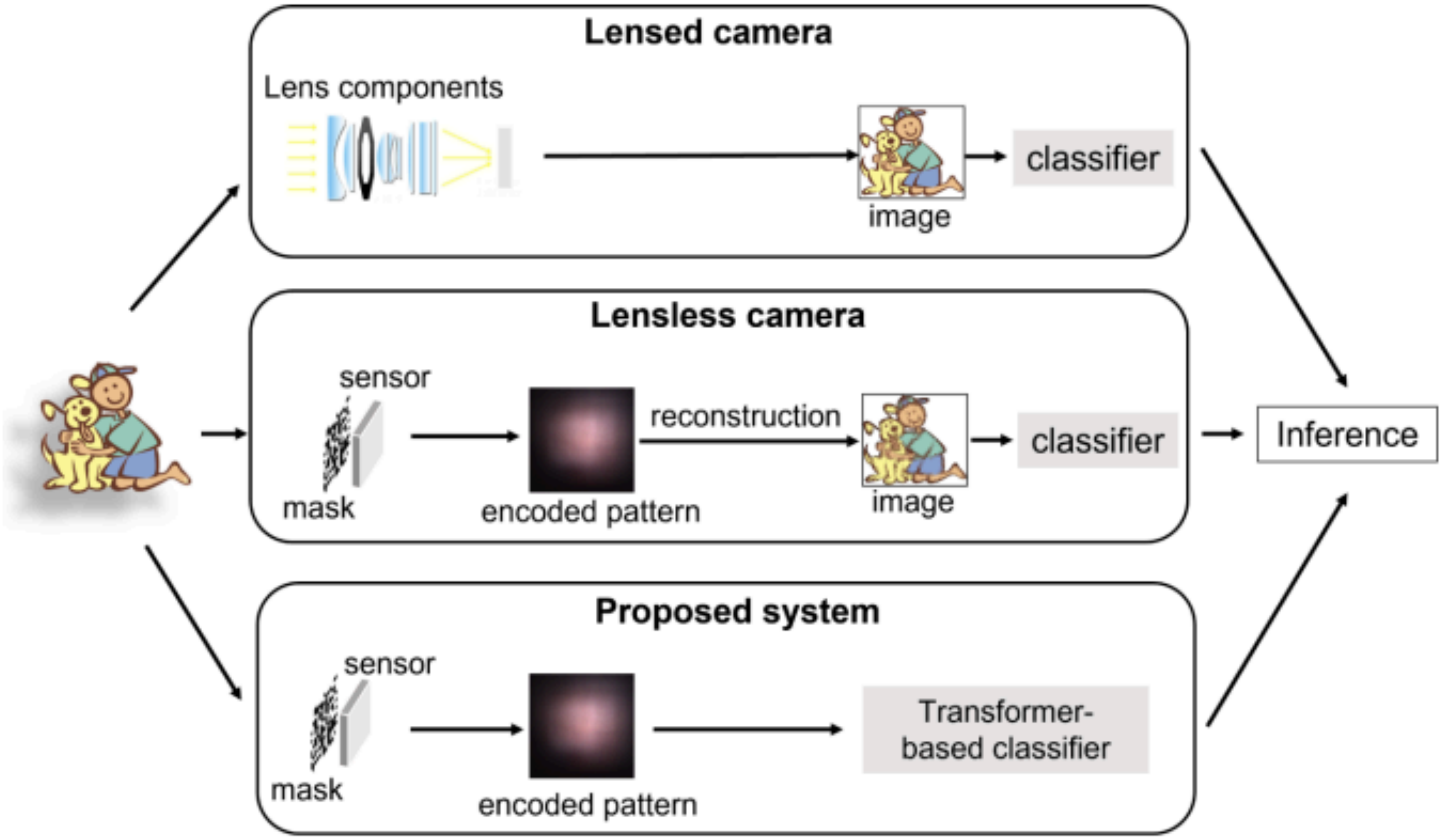
Single-shot 3D acquisition

Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica 5, 1-9 (2018)

How can AI help?



Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. Express 27, 28075-28090 (2019)



Xiuxi Pan, Xiao Chen, Tomoya Nakamura, and Masahiro Yamaguchi, "Incoherent reconstruction-free object recognition with mask-based lensless optics and the Transformer," Opt. Express 29, 37962-37978 (2021)

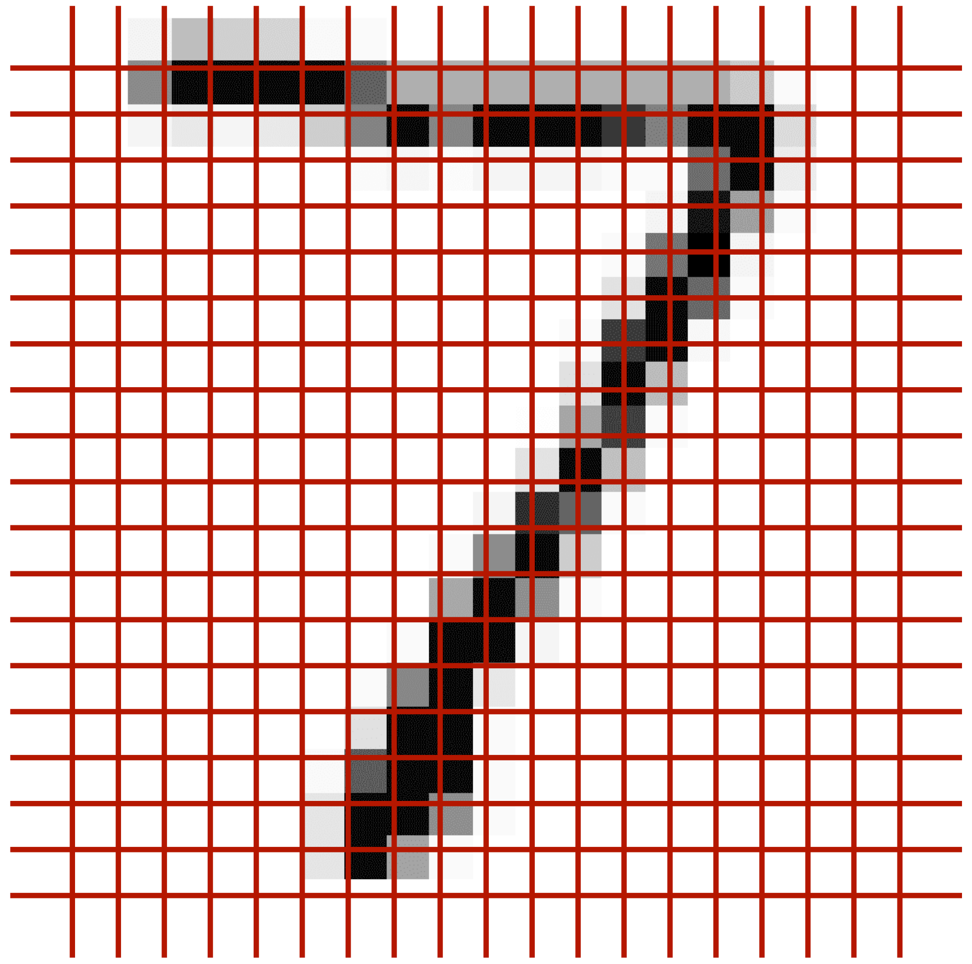
Active Learning

How can AI guide experimentation
and measurement?

How to run an experiment

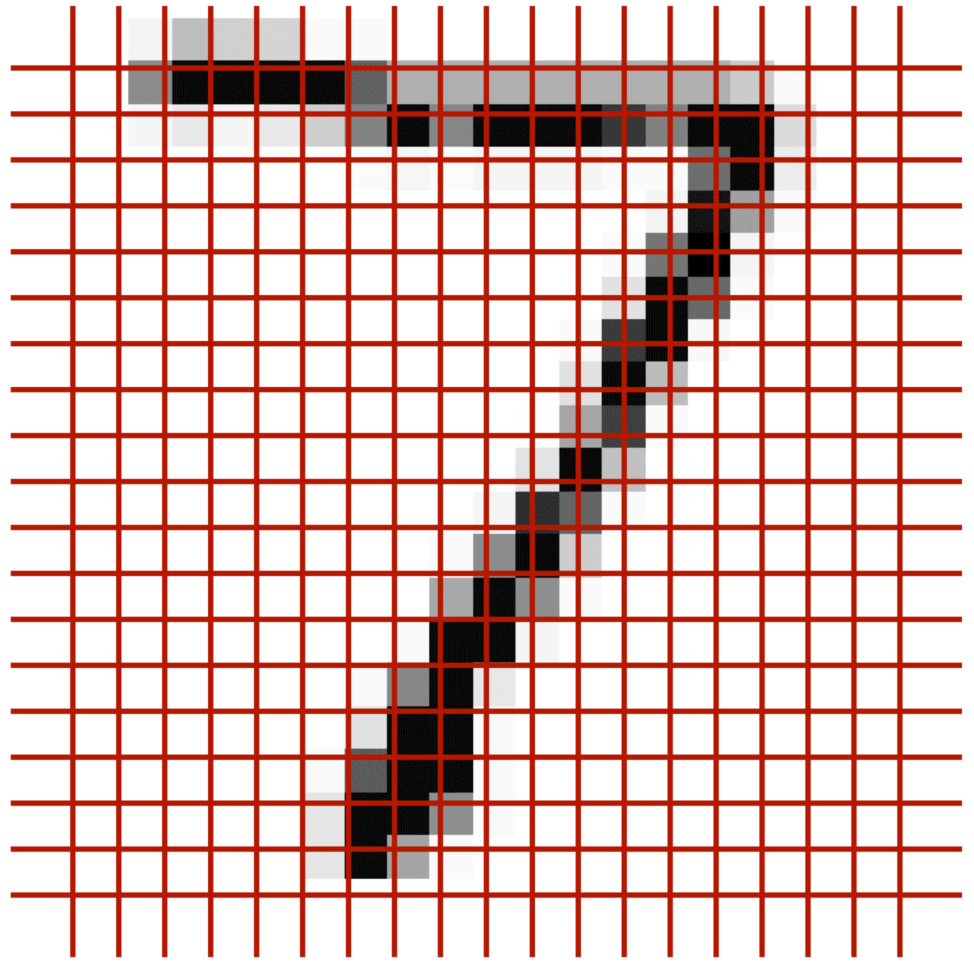
How to run an experiment

Passive Observation

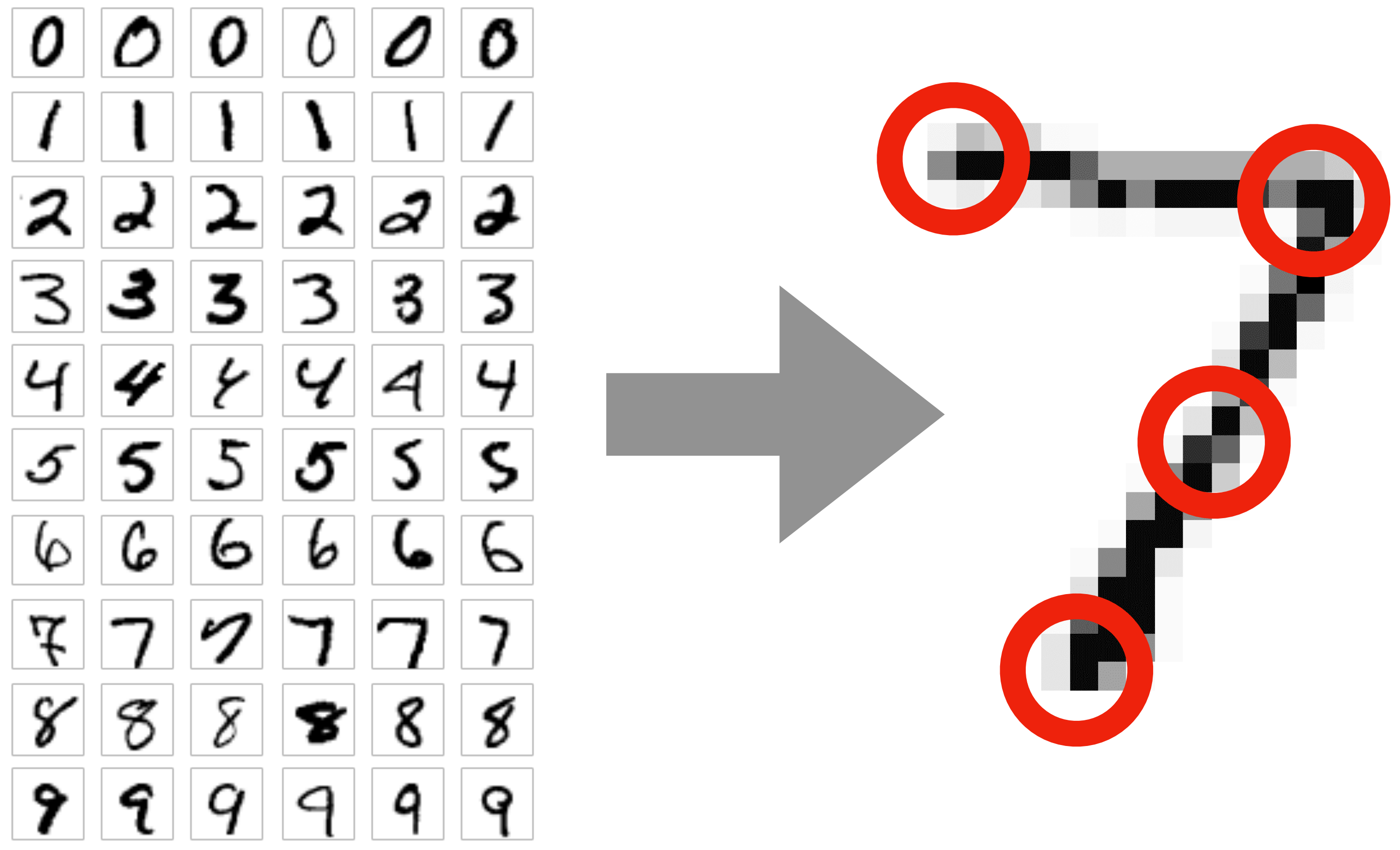


How to run an experiment

Passive Observation

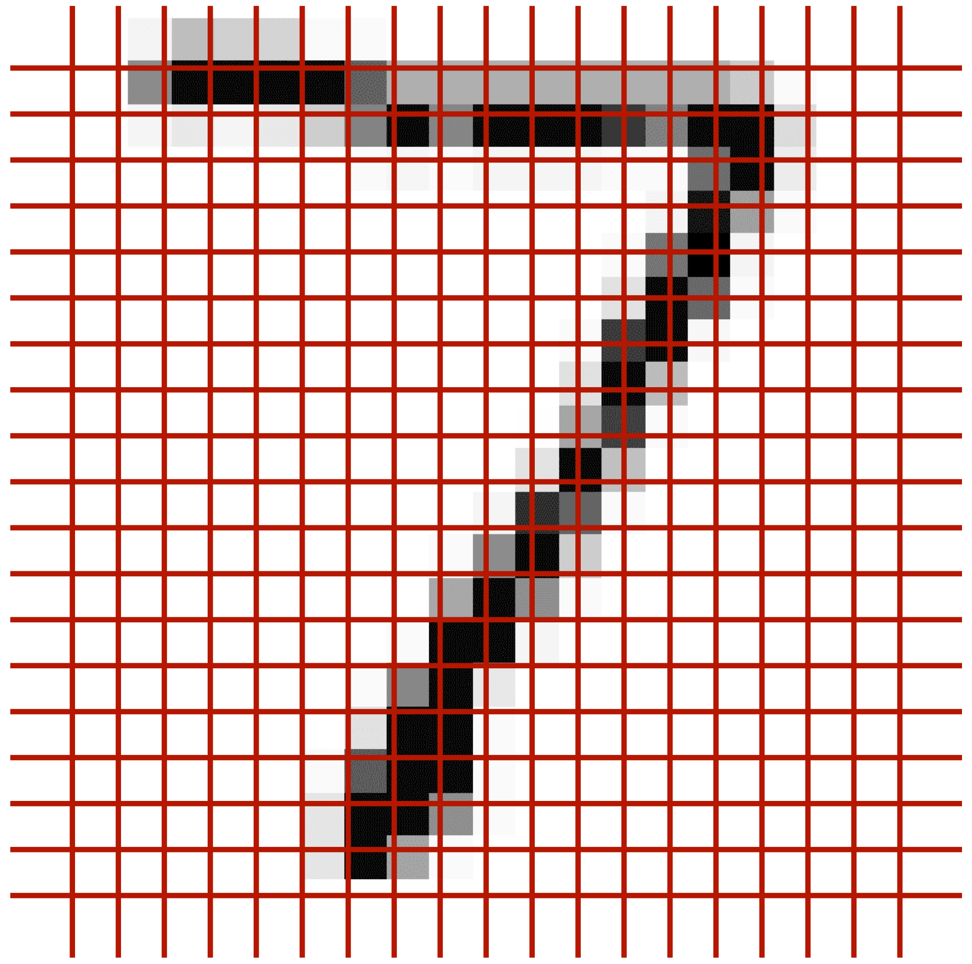


Optimal Experiment Design

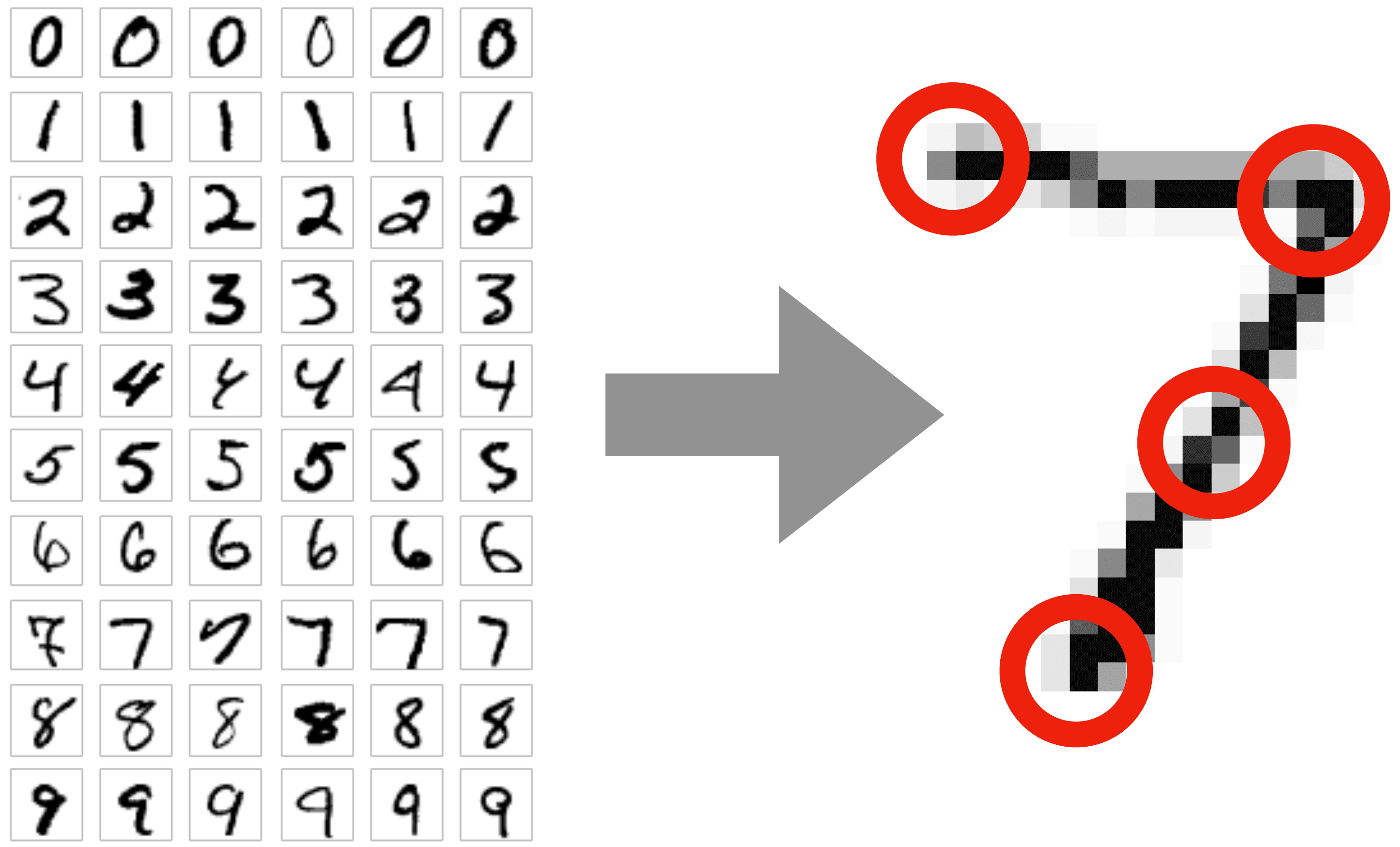


How to run an experiment

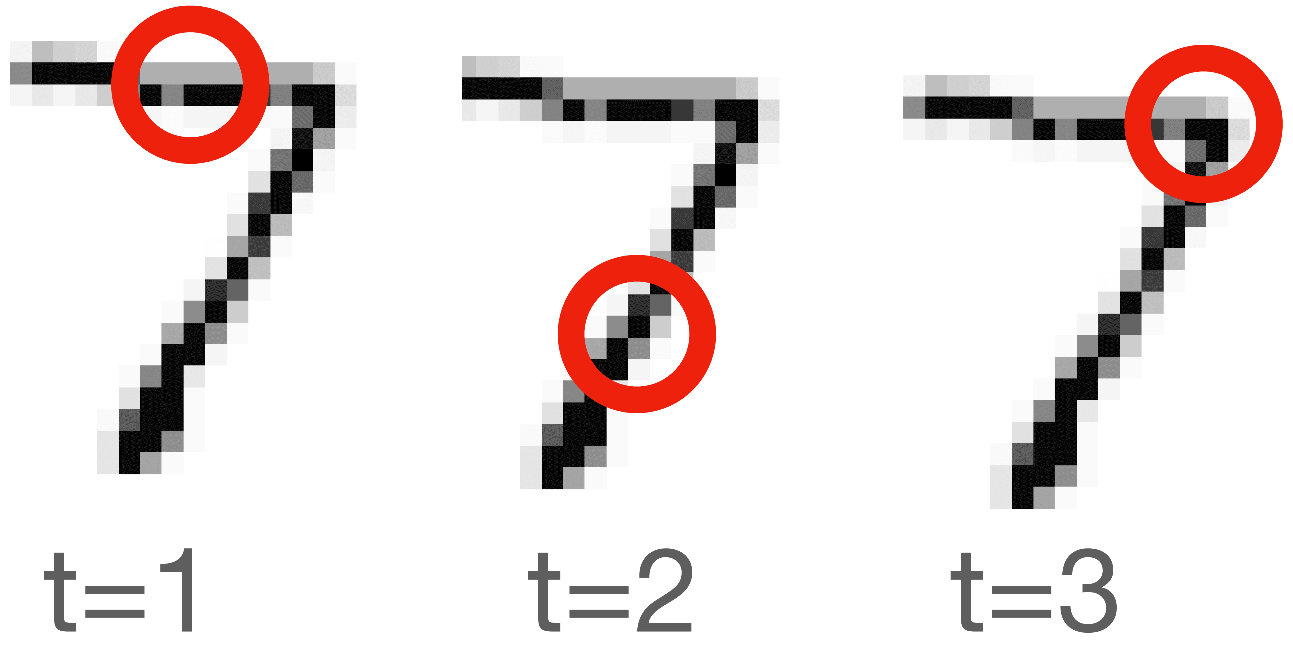
Passive Observation



Optimal Experiment Design



Active Learning

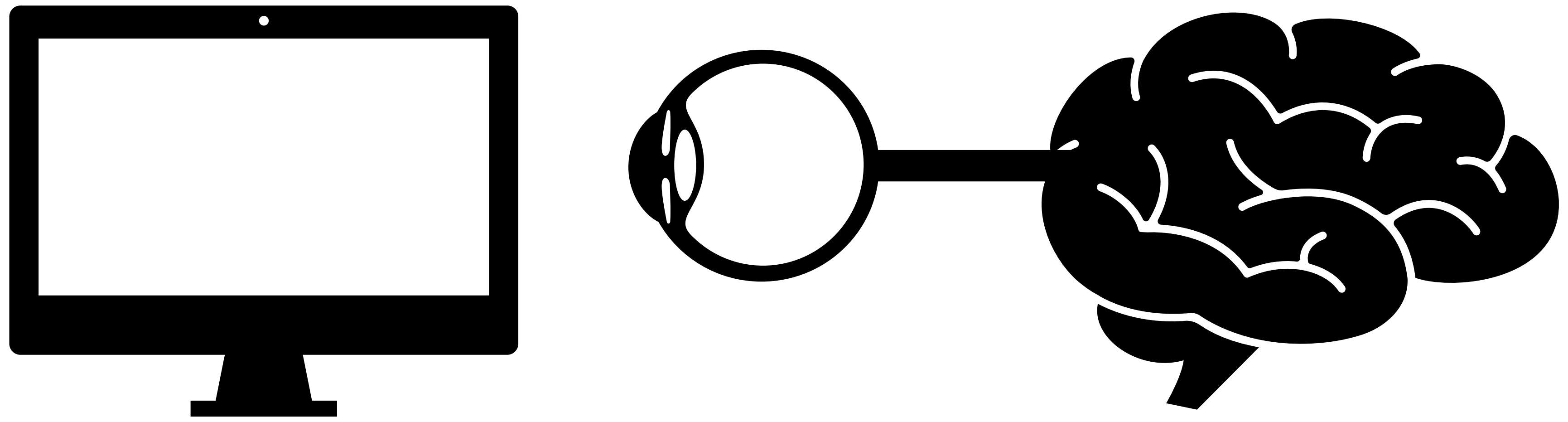


Measuring Receptive Fields

Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

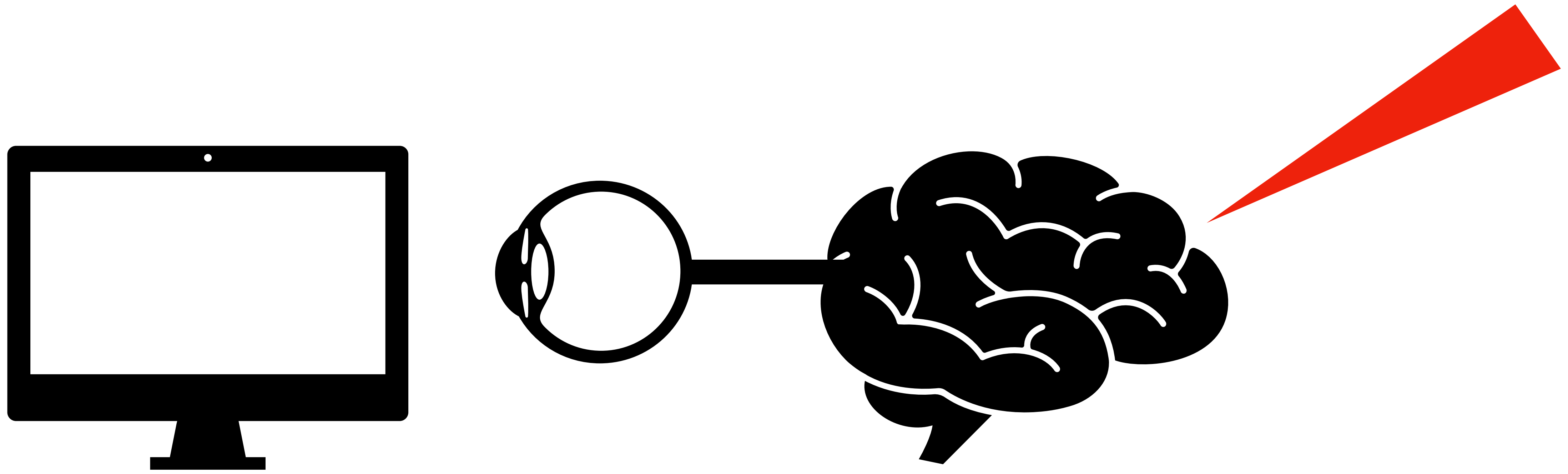
Measuring Receptive Fields



Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

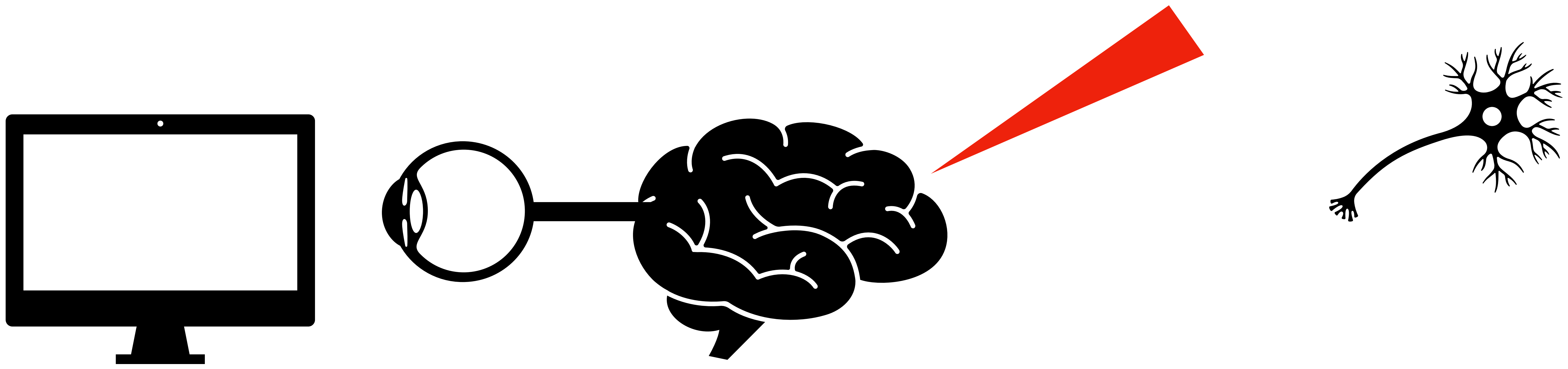
Measuring Receptive Fields



Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

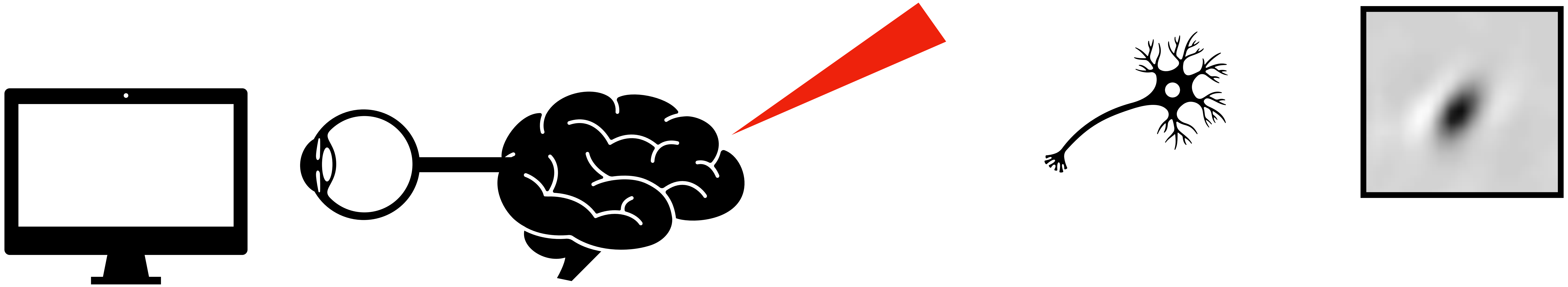
Measuring Receptive Fields



Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

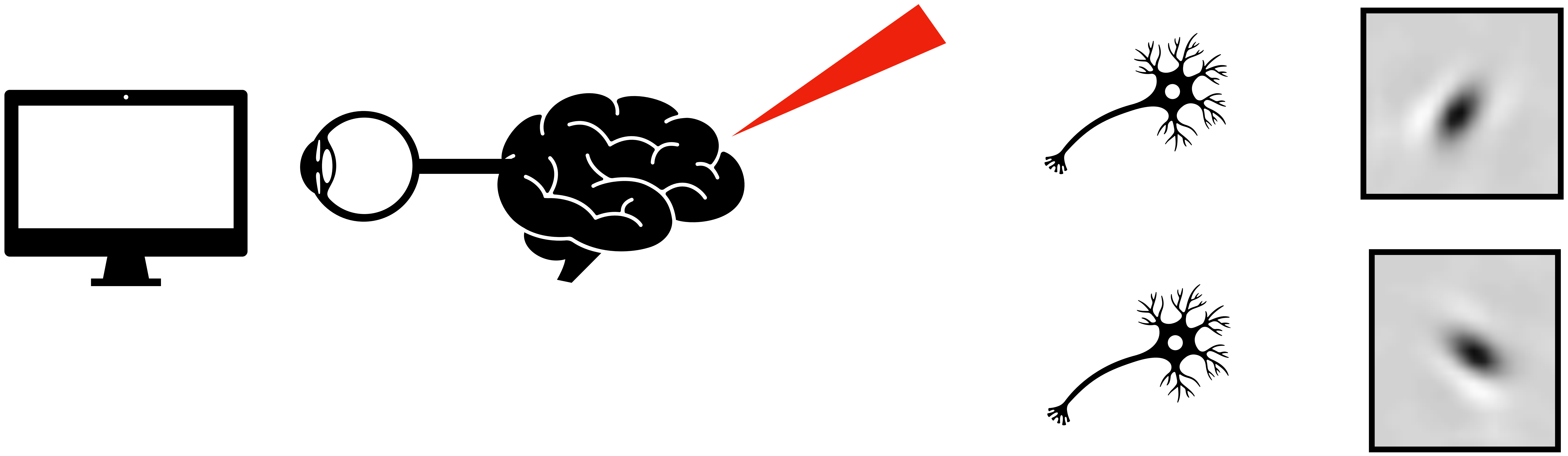
Measuring Receptive Fields



Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

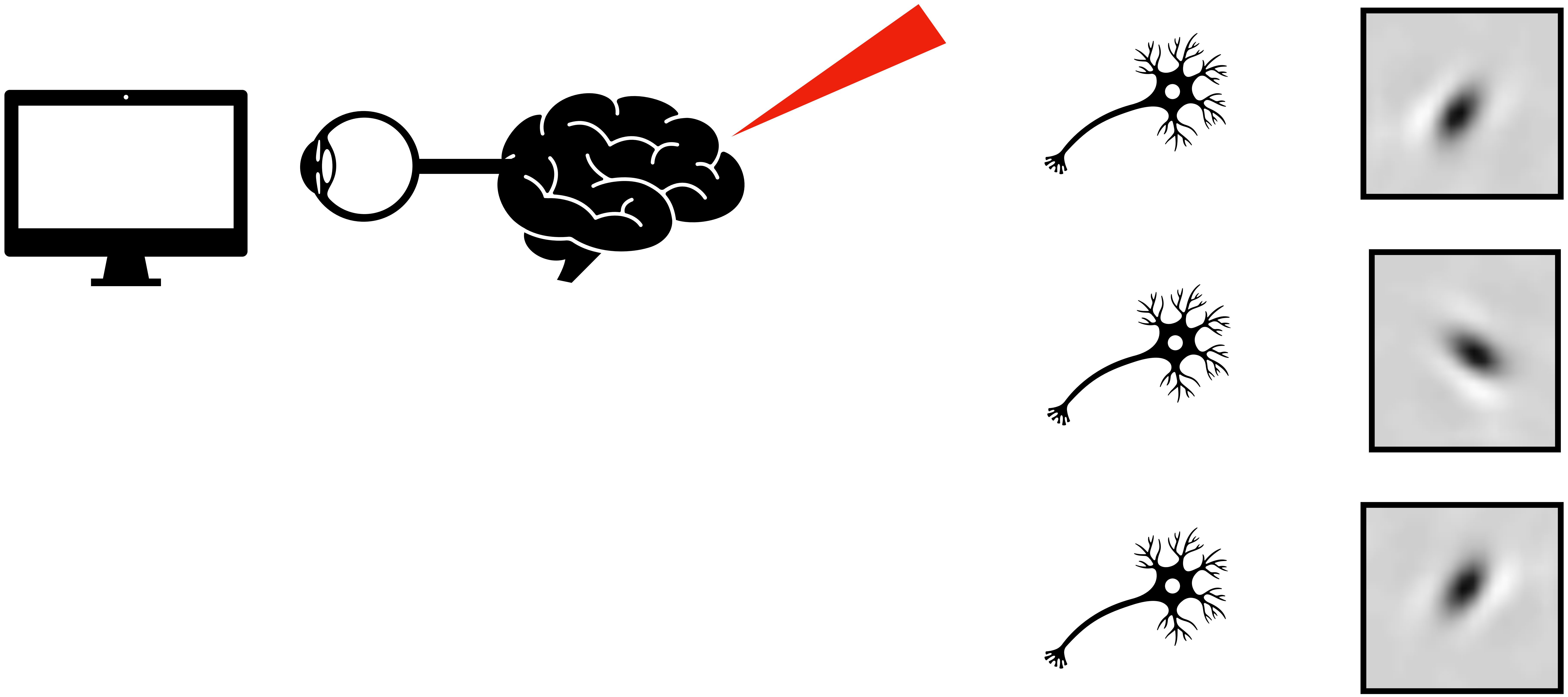
Measuring Receptive Fields



Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Measuring Receptive Fields



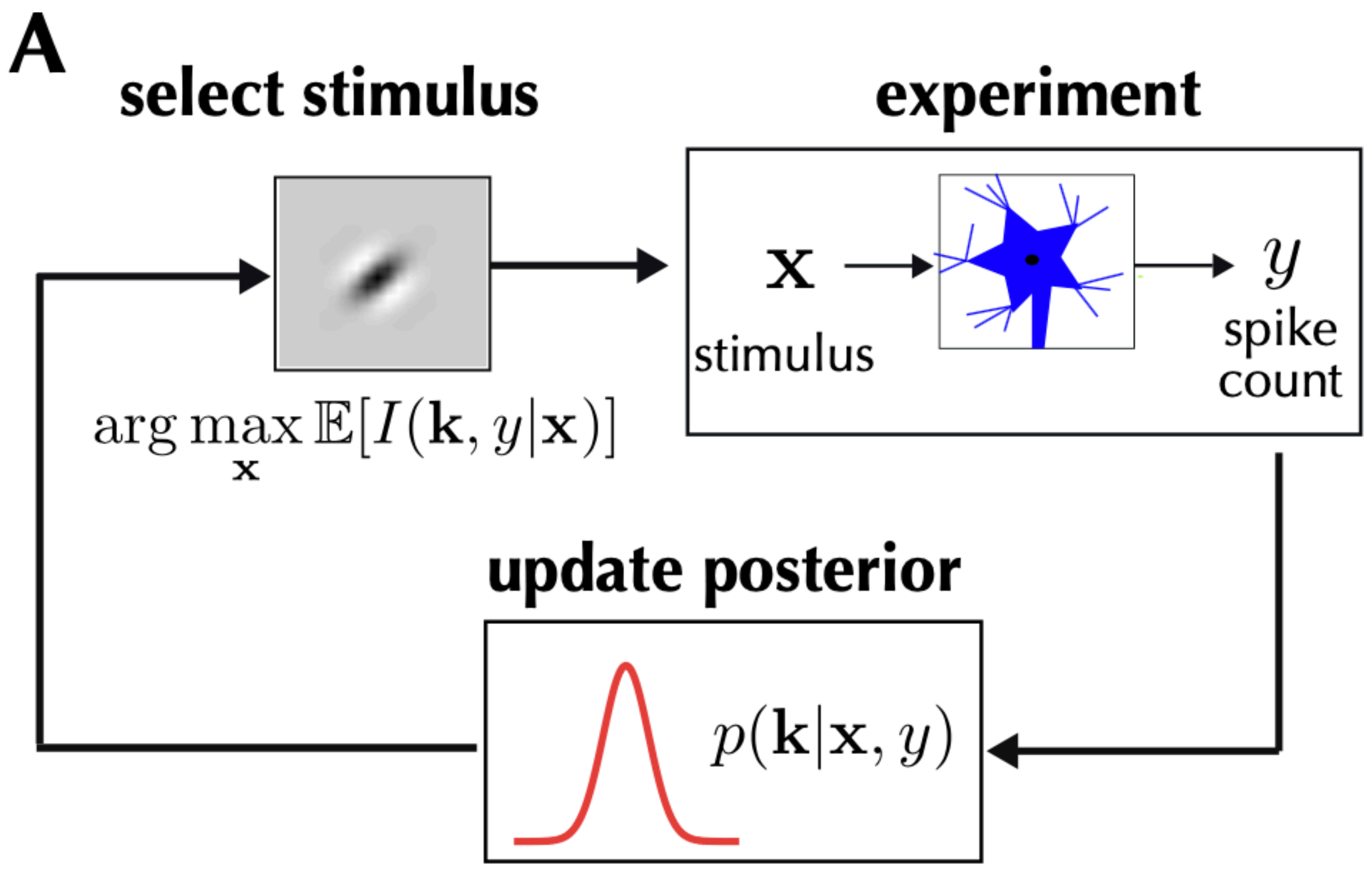
Paninski, Liam. "Asymptotic theory of information-theoretic experimental design." *Neural Computation* 17.7 (2005): 1480-1507.

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Active Learning Receptive Fields

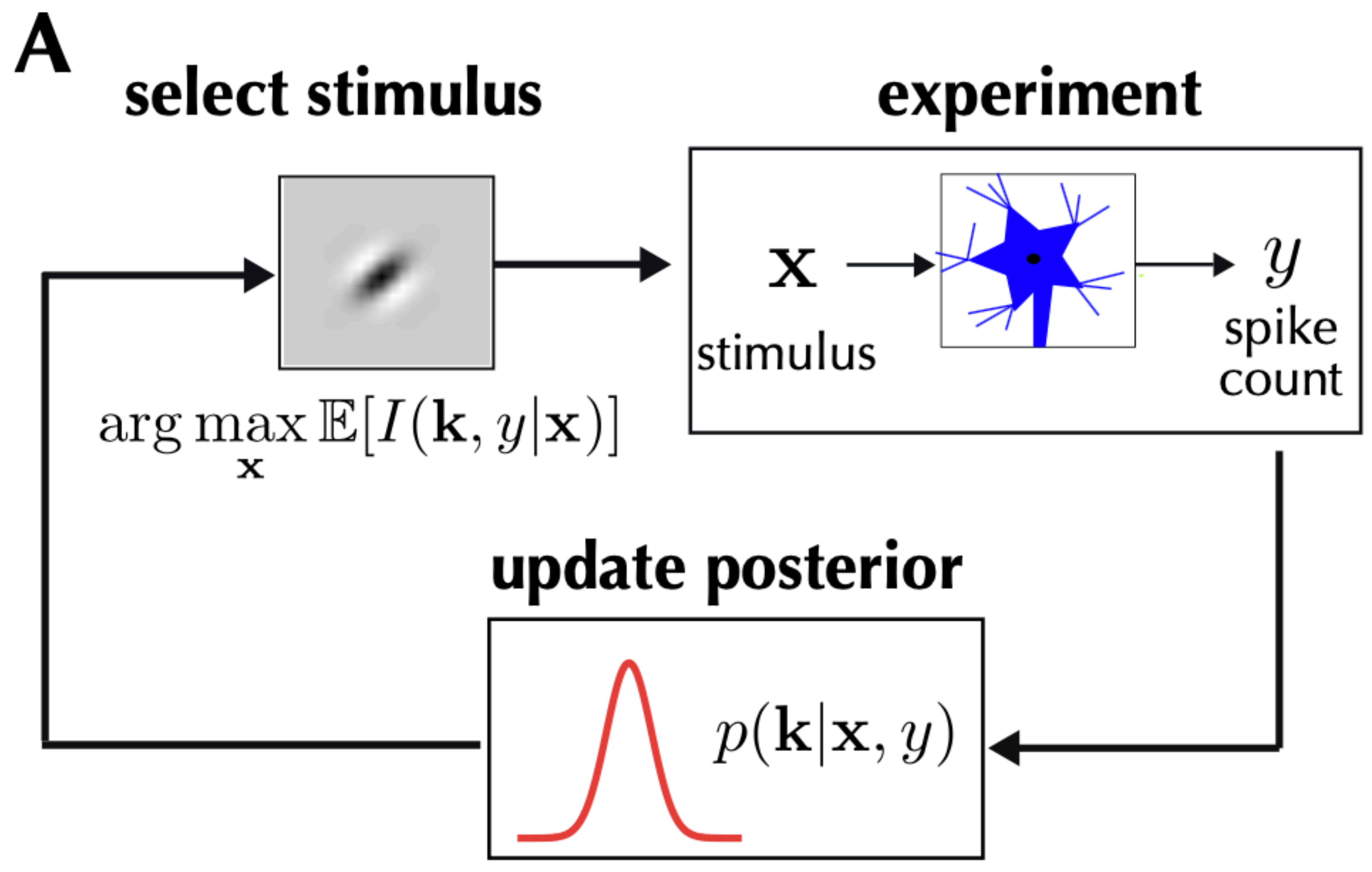
Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

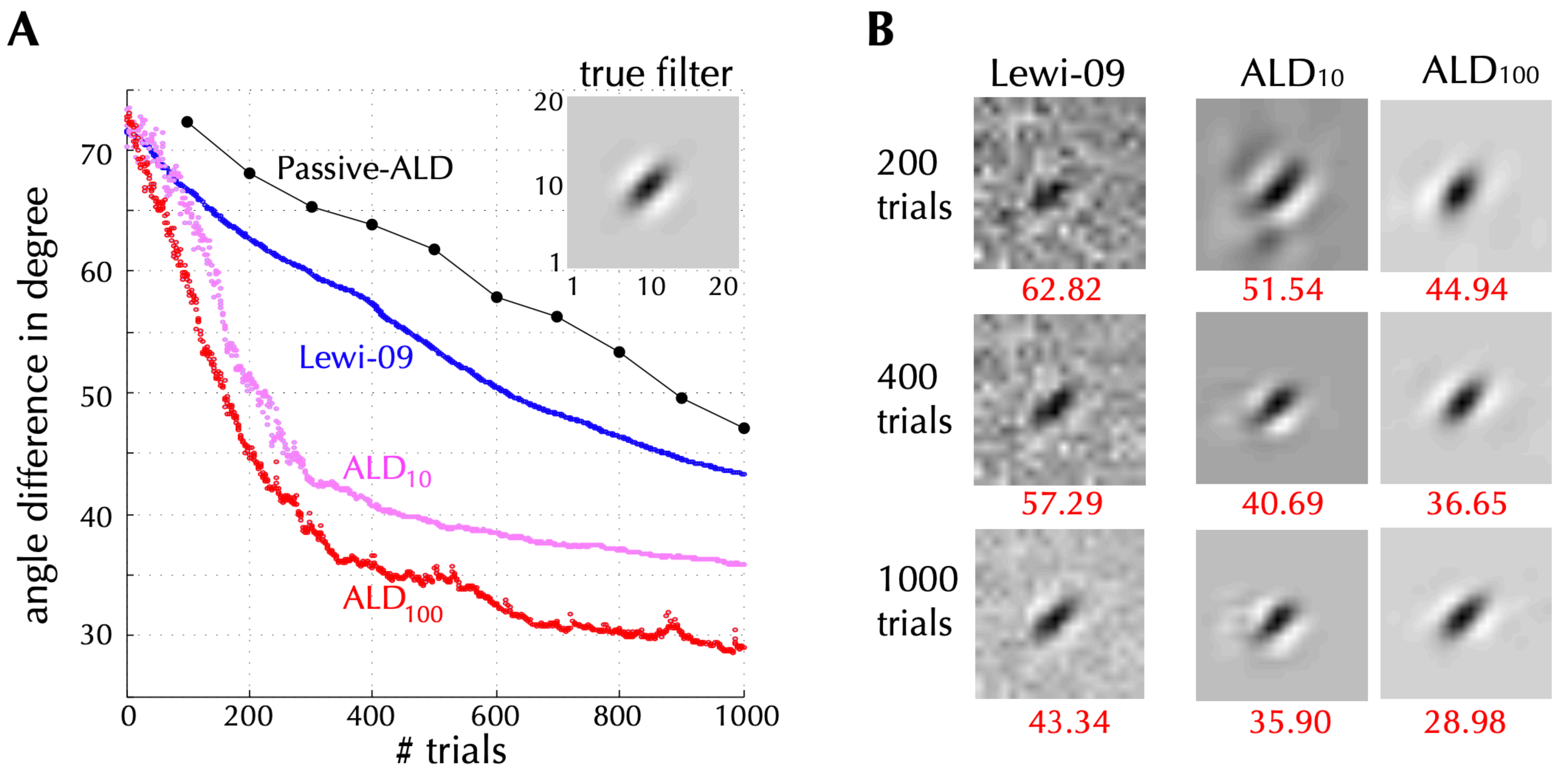
Active Learning Receptive Fields



Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

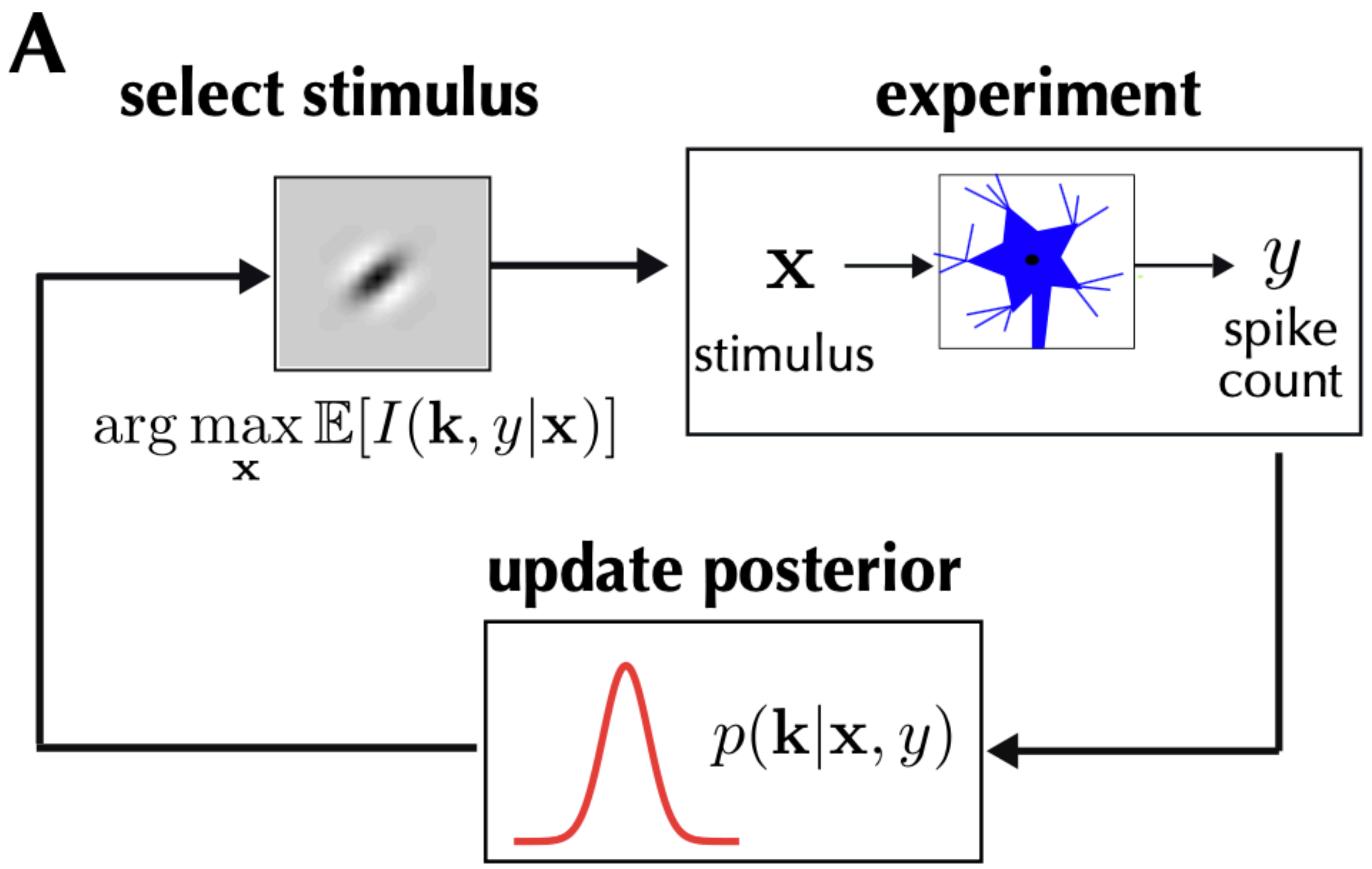
Active Learning Receptive Fields

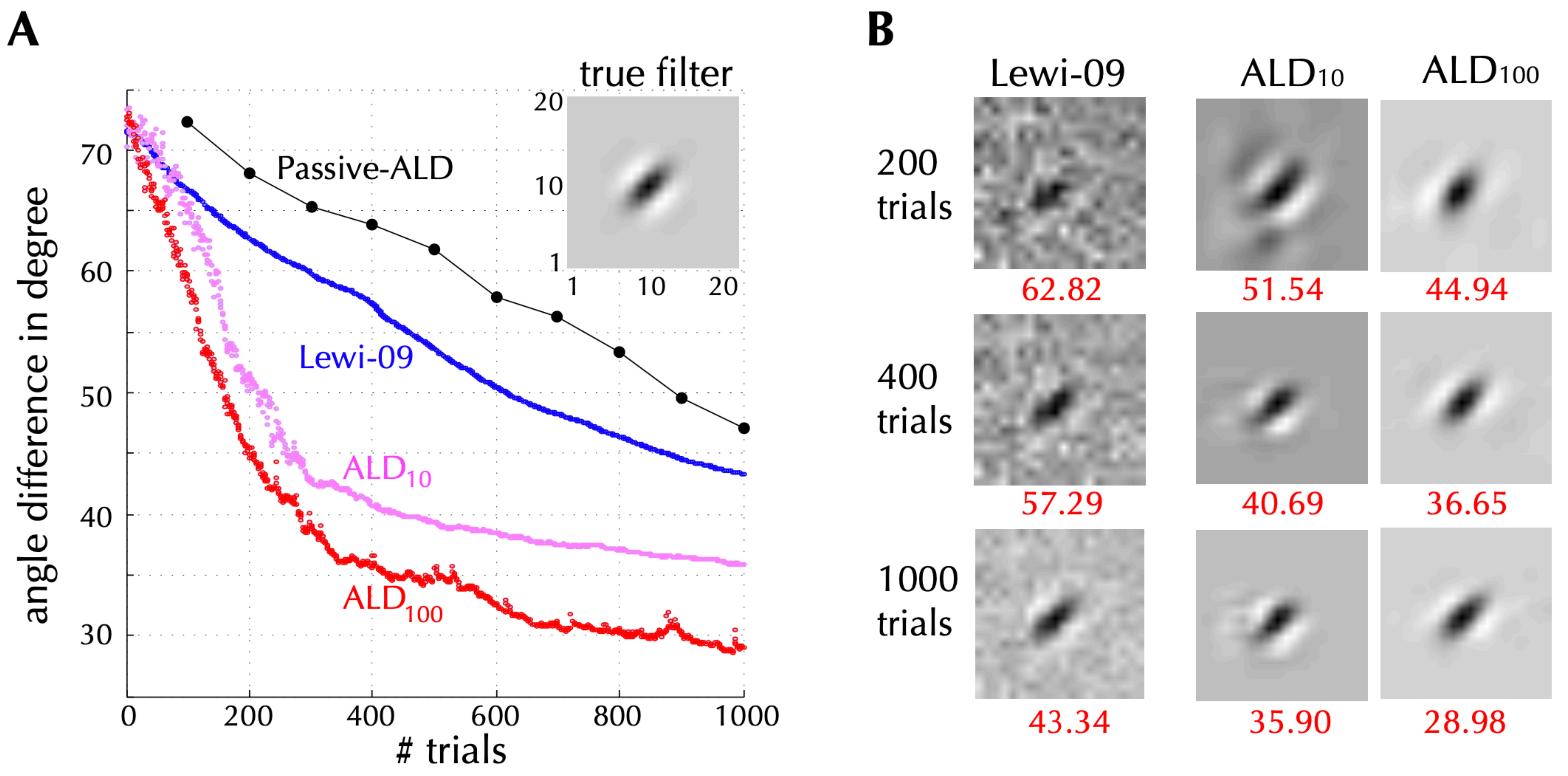




Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Active Learning Receptive Fields

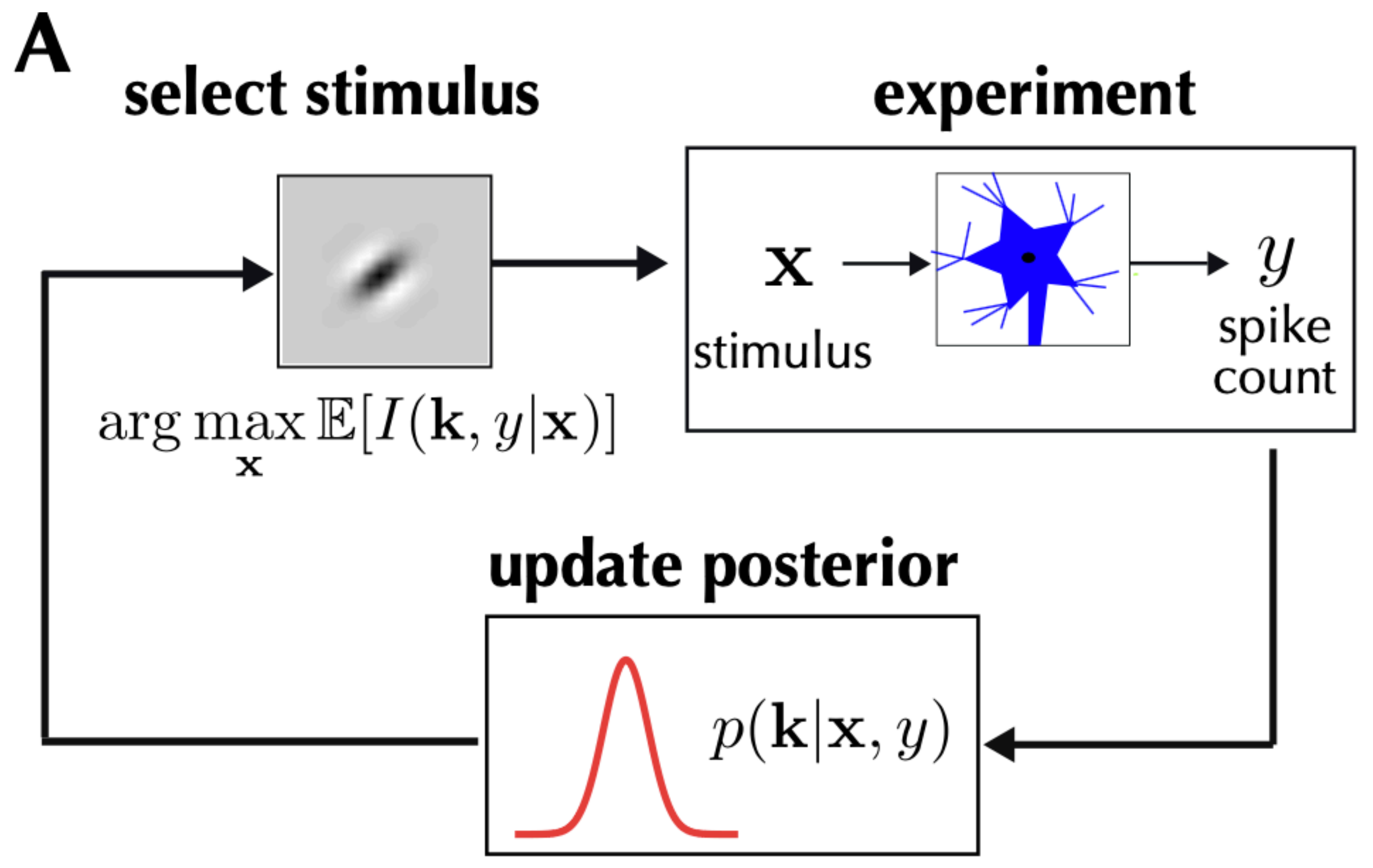




So why isn't this used?

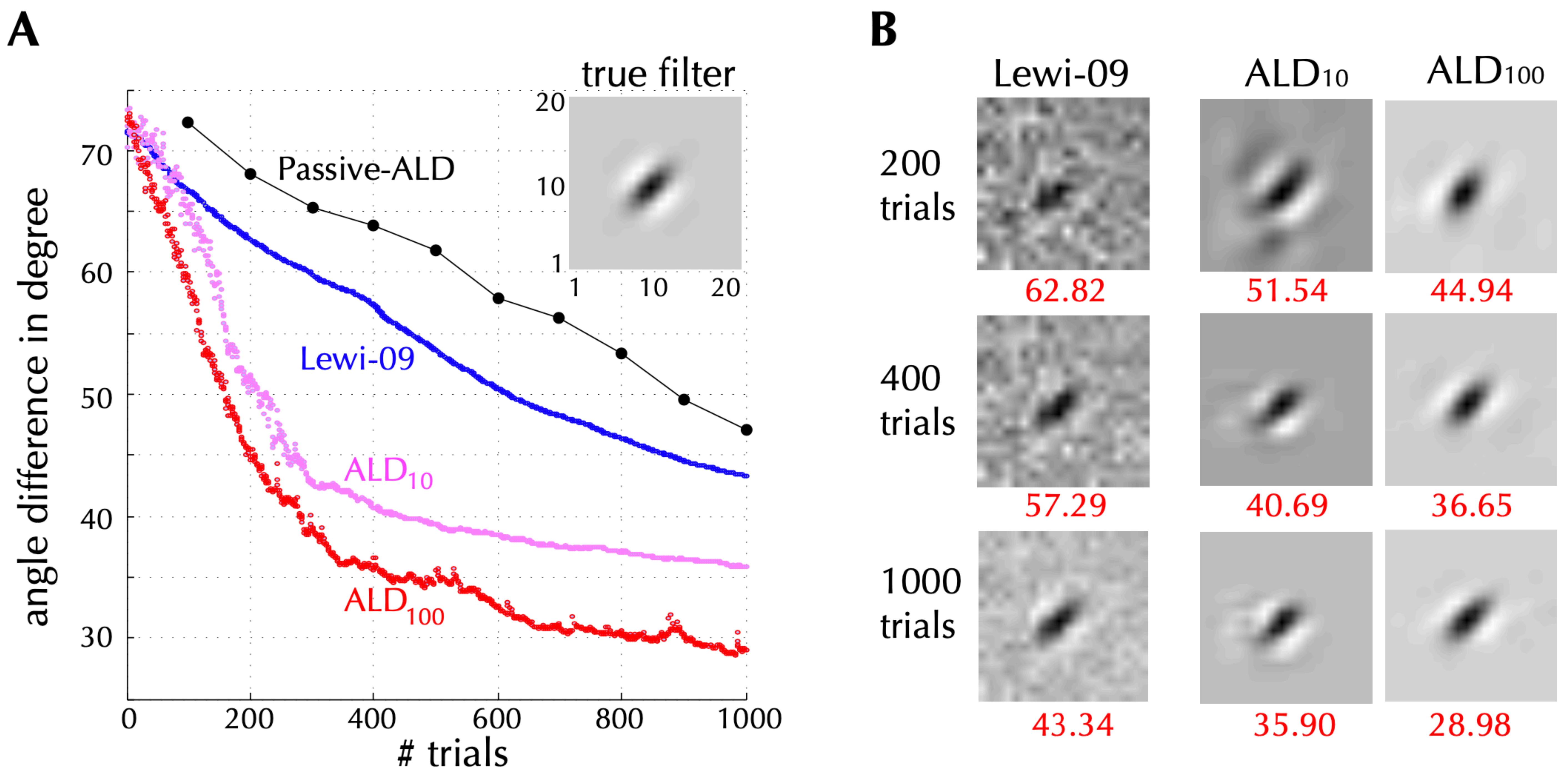
Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Active Learning Receptive Fields



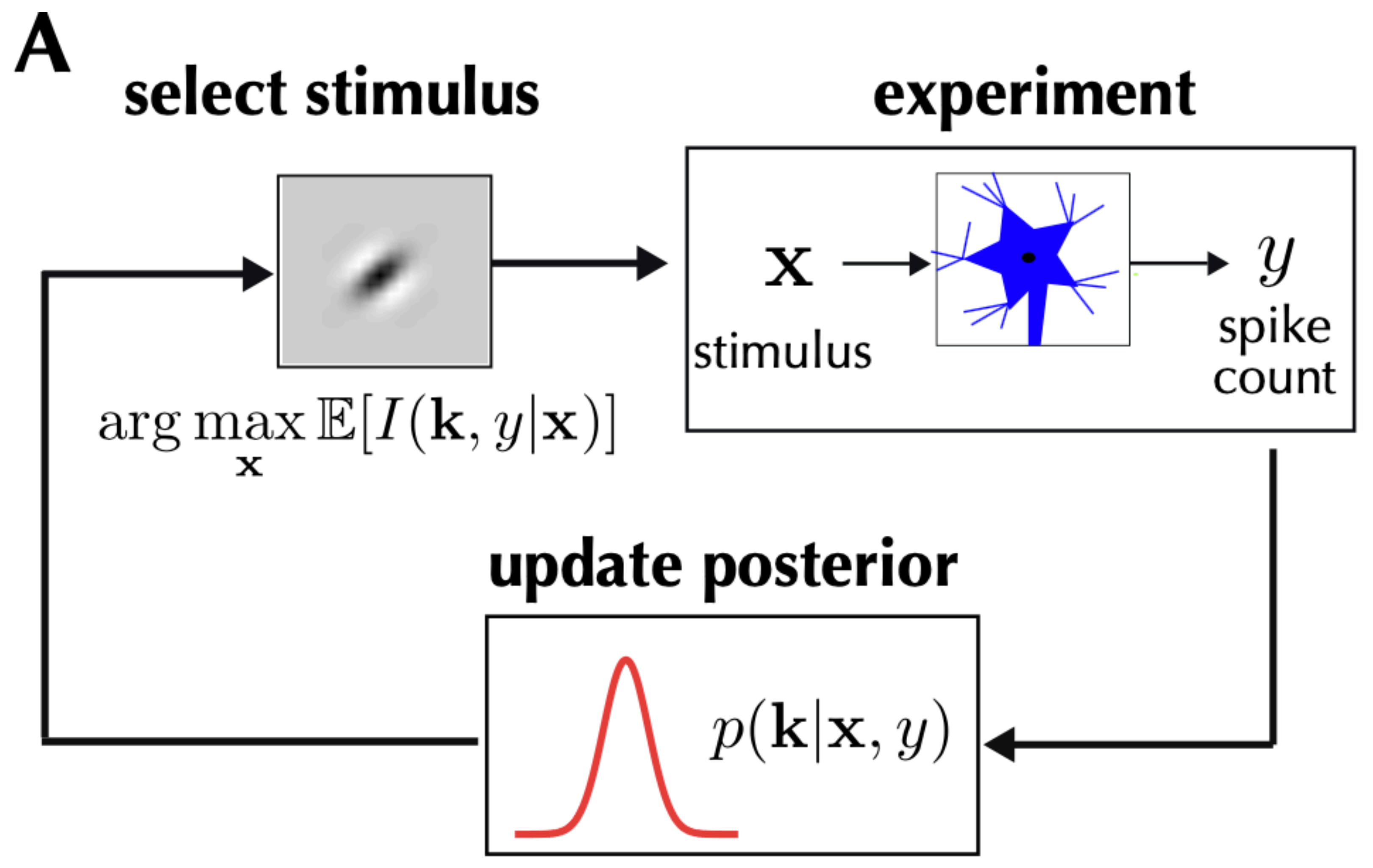
So why isn't this used?

- Implementation is hard and complex!



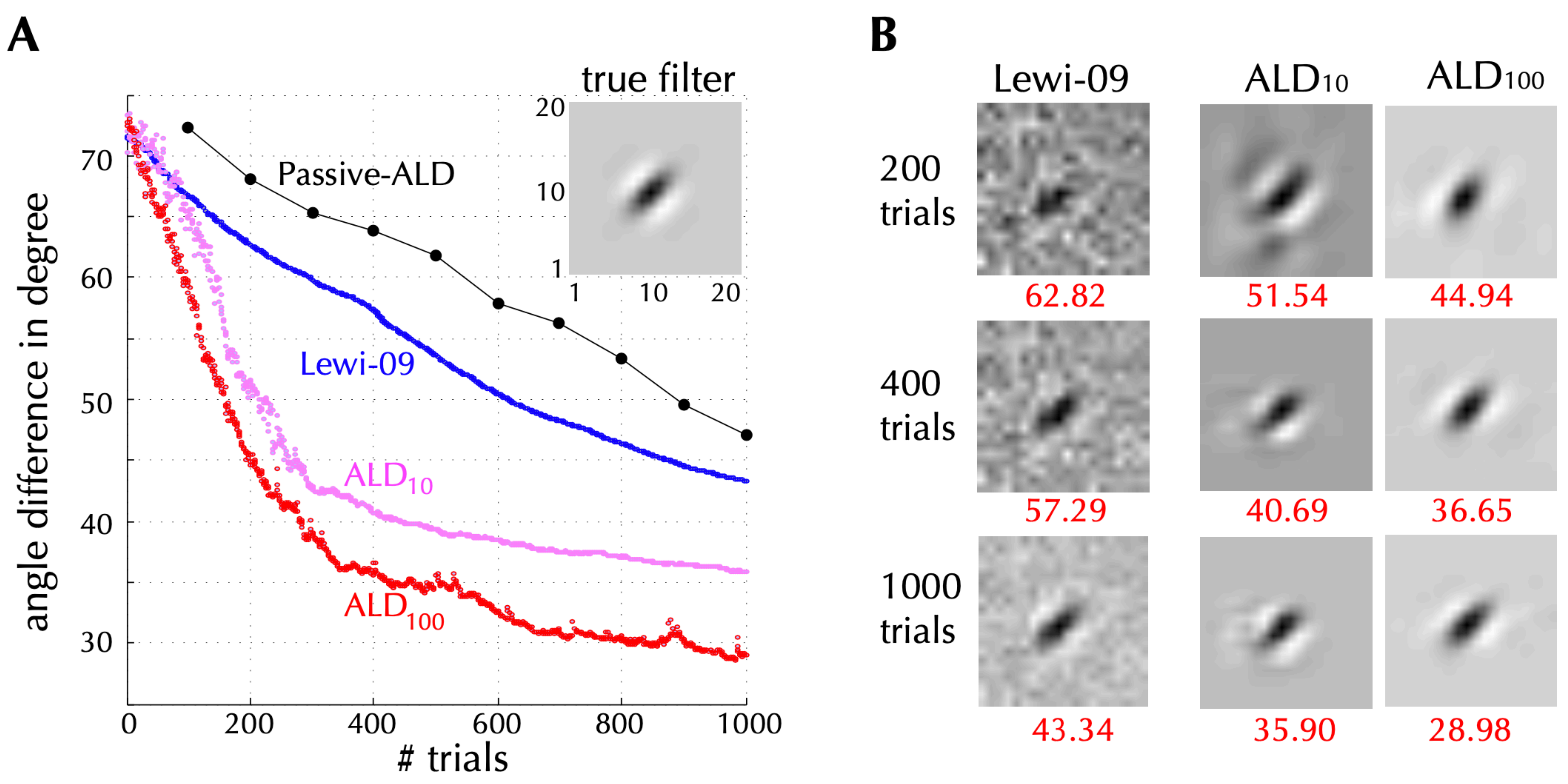
Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Active Learning Receptive Fields



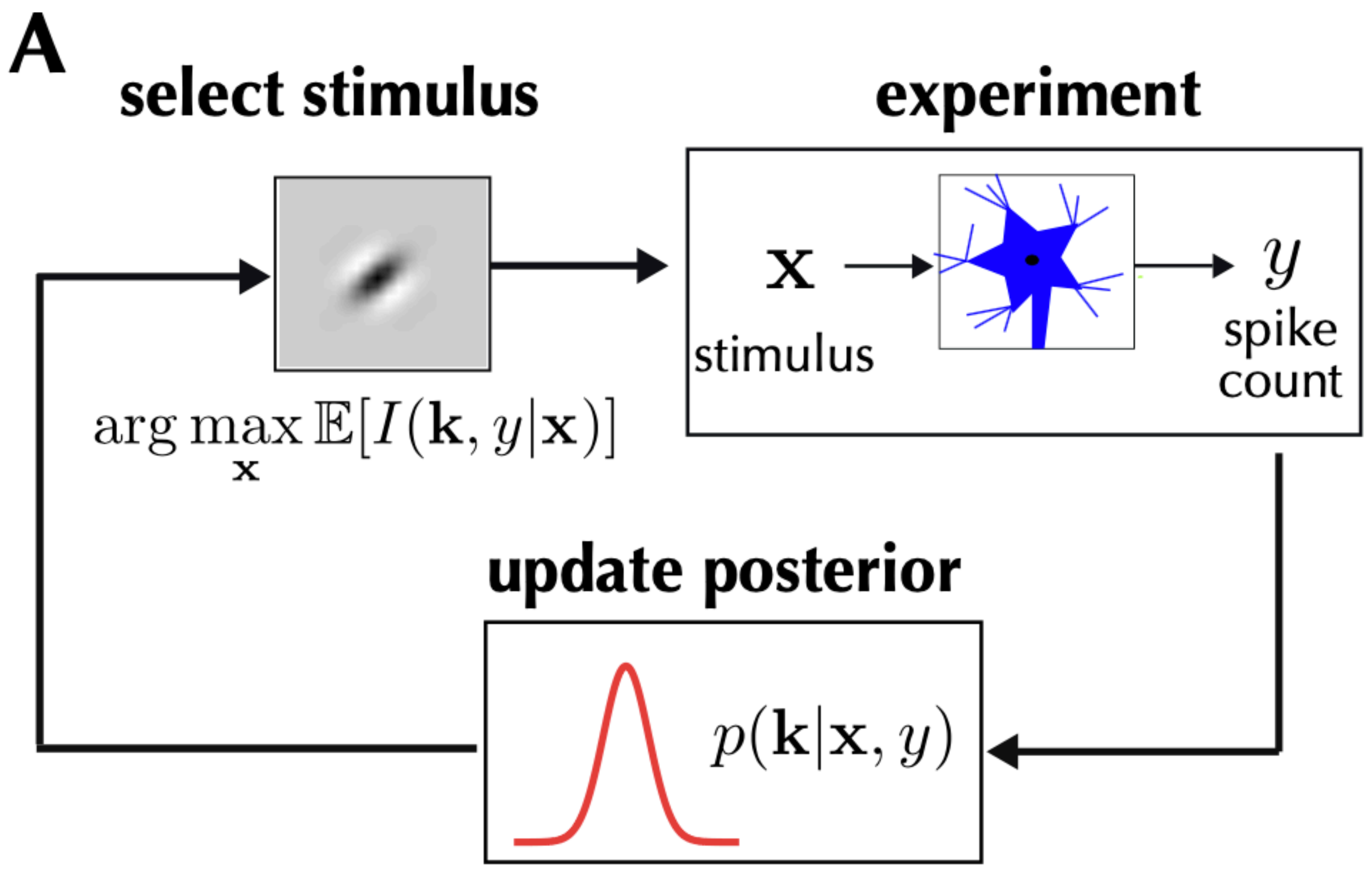
So why isn’t this used?

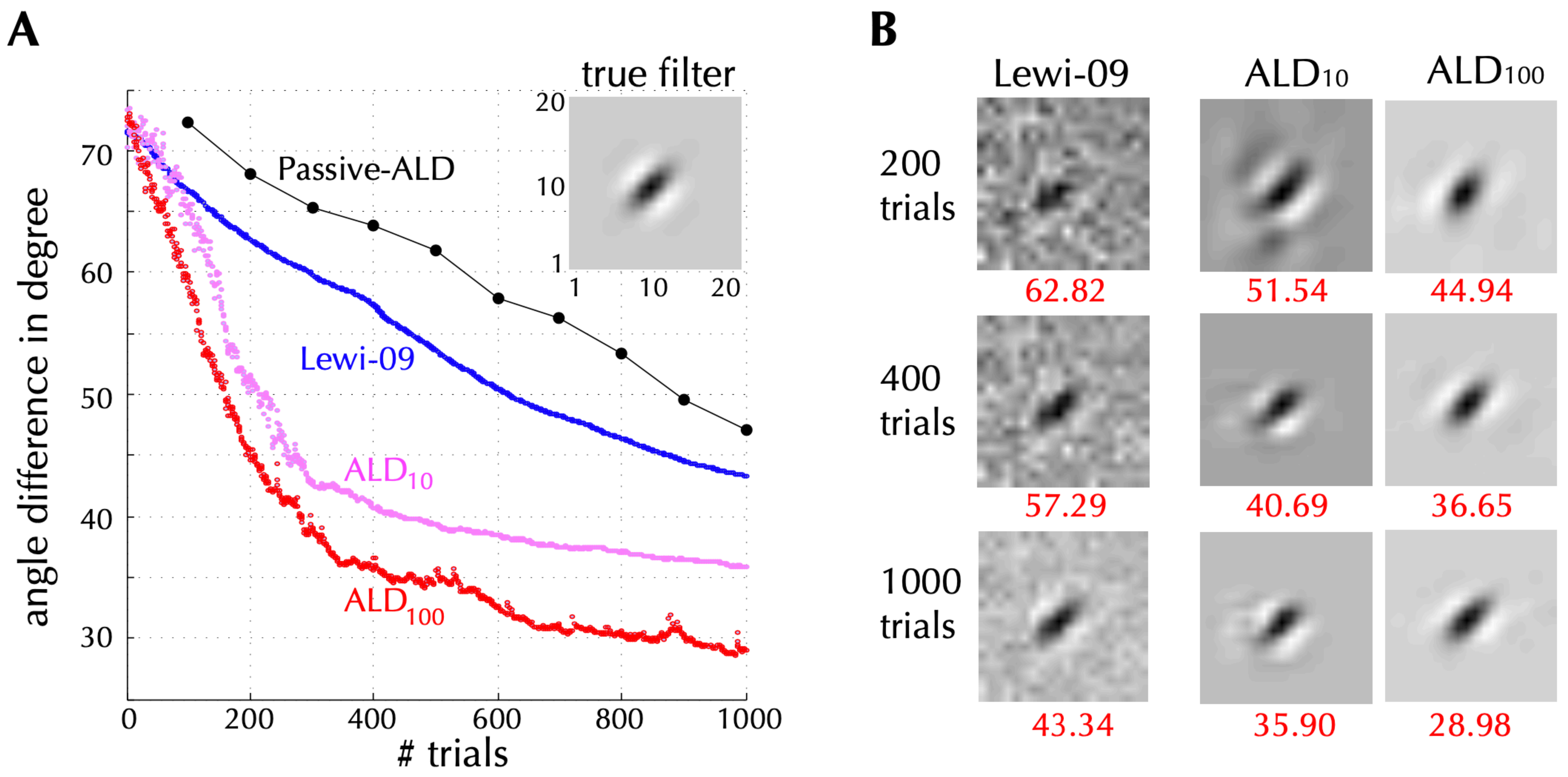
- Implementation is hard and complex!
- Even though theory bounds error, theory is based on a model (which may be wrong!)



Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Active Learning Receptive Fields





So why isn't this used?

- Implementation is hard and complex!
- Even though theory bounds error, theory is based on a model (which may be wrong!)
- Easier to just spend $3B and scale up experiments

Bayesian Active Learning with localized priors for fast receptive field characterization. Park, Pillow NeurIPS 2012

Finding the sweet spot

How much AI progress is necessary?

How much AI progress is necessary?

How much AI progress is necessary?

“How can I analyze my data?”

How much AI progress is necessary?

How much AI progress is necessary?

“Existing algorithms aren’t quite right”

“How can I analyze my data?”

“Can you replace my grad student?”

How much AI progress is necessary?

“Existing algorithms aren’t quite right”

Regression

Sentience

“How can I analyze my data?”

“Can you replace my grad student?”

How much AI progress is necessary?

“Existing algorithms aren’t quite right”

Regression

Sentience

“How can I analyze my data?”

sweet spot for us

“Can you replace my grad student?”

How much AI progress is necessary?

“Existing algorithms aren’t quite right”

Regression

Sentience

“How can I analyze my data?”

sweet spot for us

- ML provides fundamentally new capabilities but is “mostly there” already

“Can you replace my grad student?”

How much AI progress is necessary?

“Existing algorithms aren’t quite right”

Regression

Sentience

“How can I analyze my data?”

sweet spot for us

- ML provides fundamentally new capabilities but is “mostly there” already
- Creative ideas beyond existing work — rethinking what’s possible

“Can you replace my grad student?”

How much AI progress is necessary?

"Existing algorithms aren't quite right"

Regression

Sentience

"How can I analyze my data?"

sweet spot for us

- ML provides fundamentally new capabilities but is "mostly there" already
- Creative ideas beyond existing work — rethinking what's possible
- Existing baselines so you understand how much progress can be made

"Can you replace my grad student?"

What is a robot?

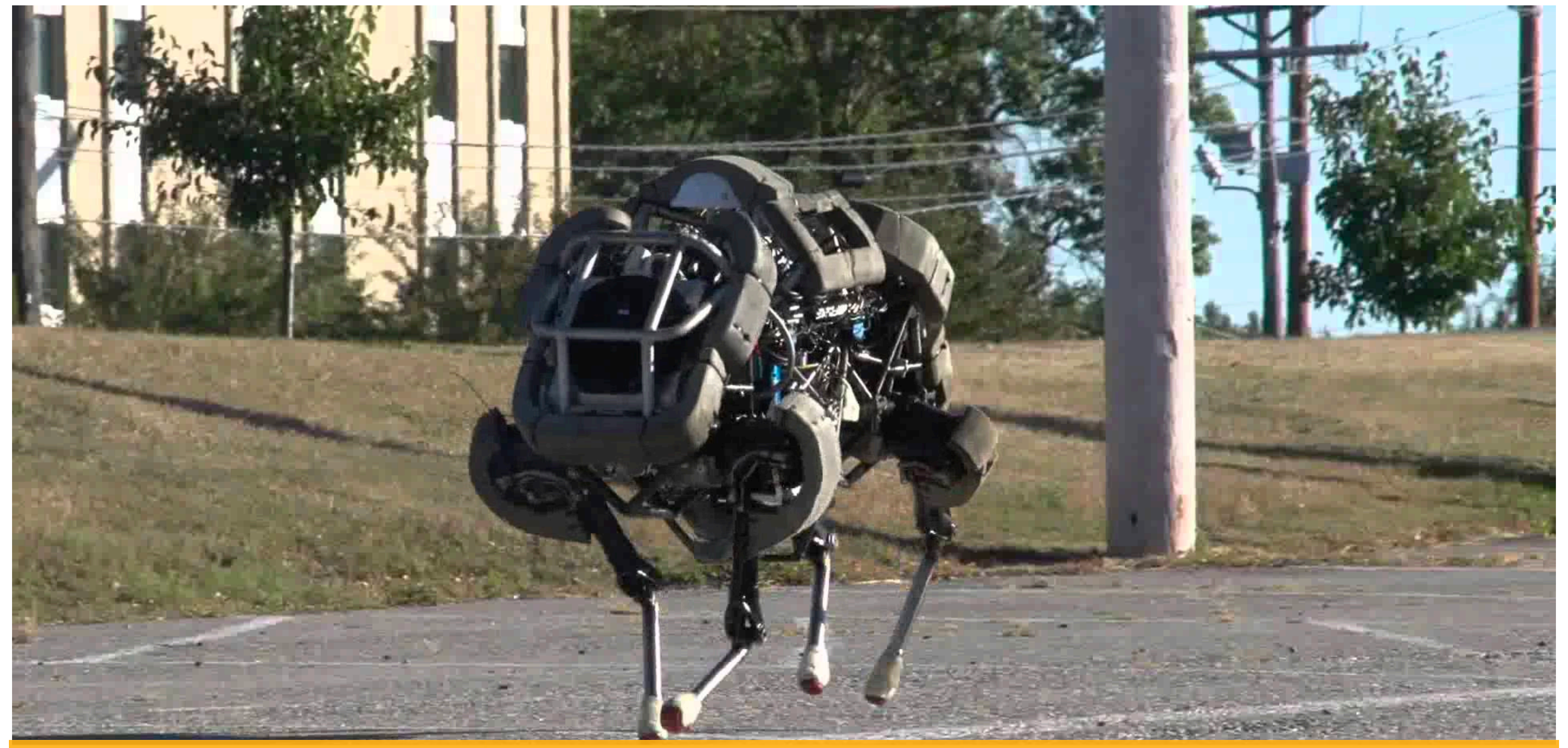
What is a robot?

What is a robot?

What is a robot?


T-800

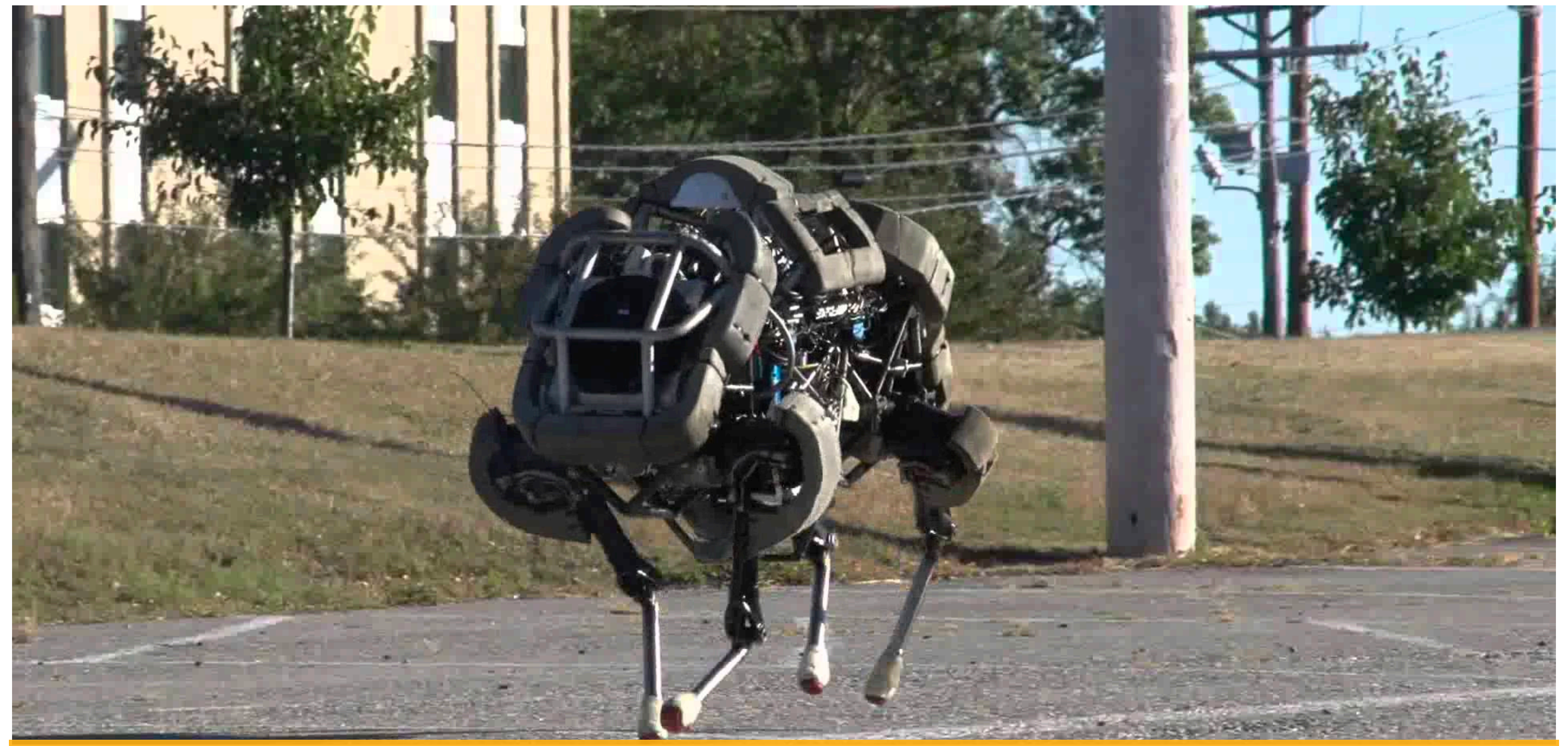

KIVA WAREHOUSE ROBOT


BIG DOG FROM BOSTON D.

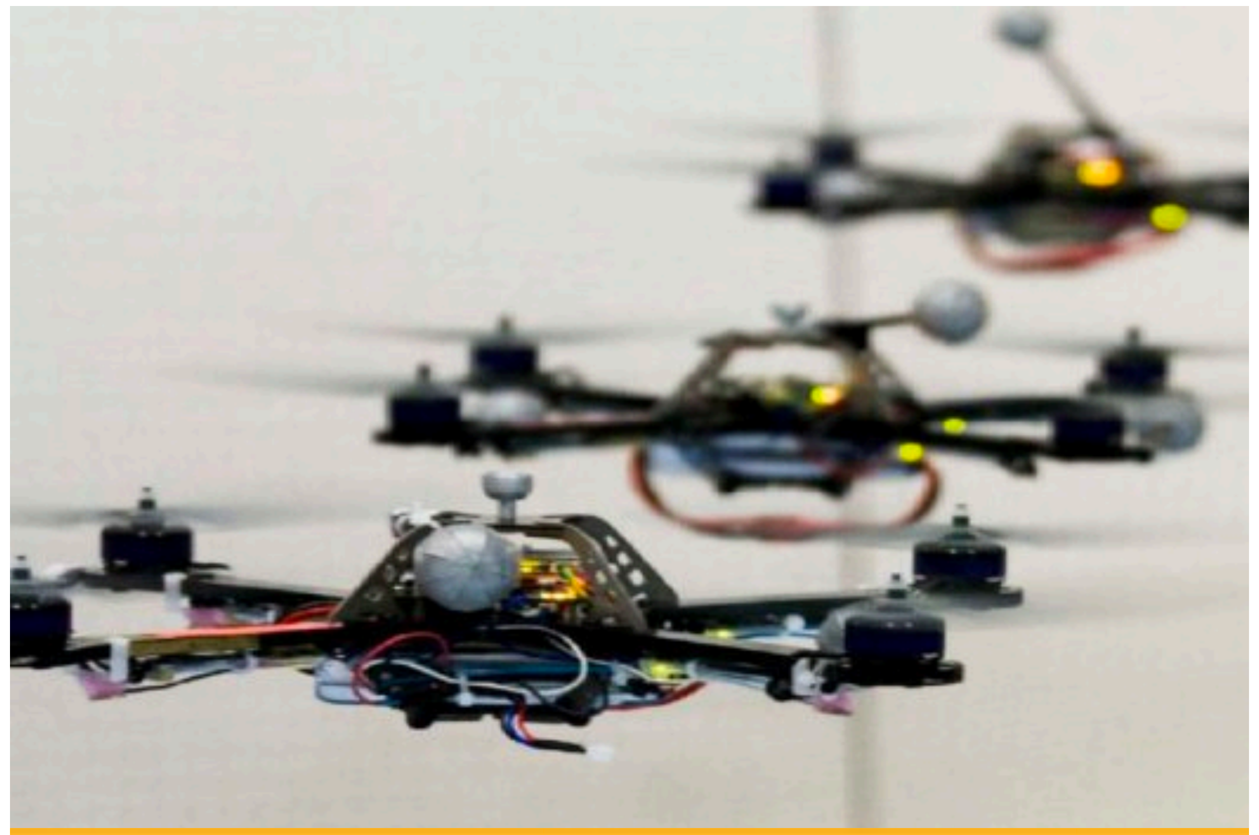
What is a robot?

T-800

KIVA WAREHOUSE ROBOT



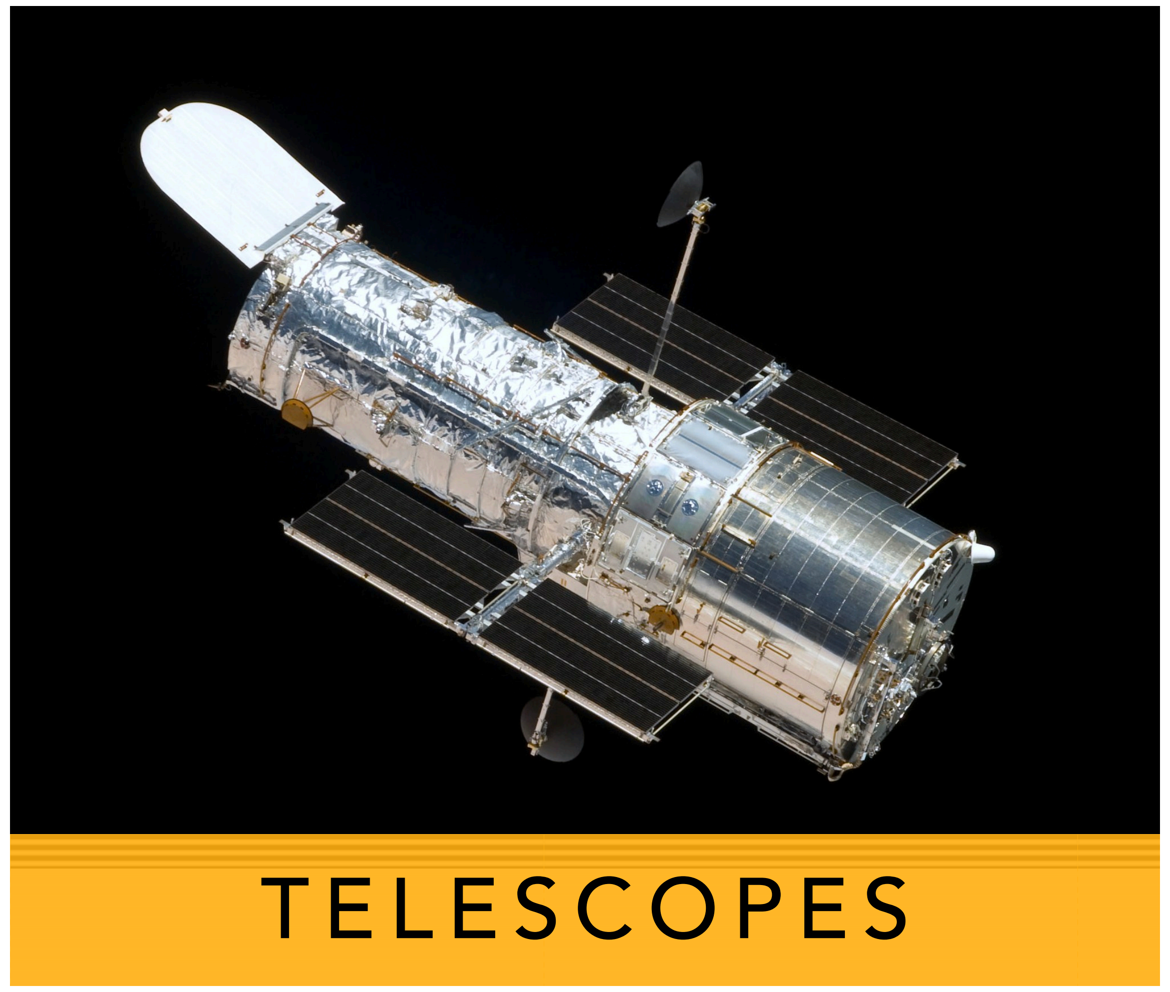
BIG DOG FROM BOSTON D.



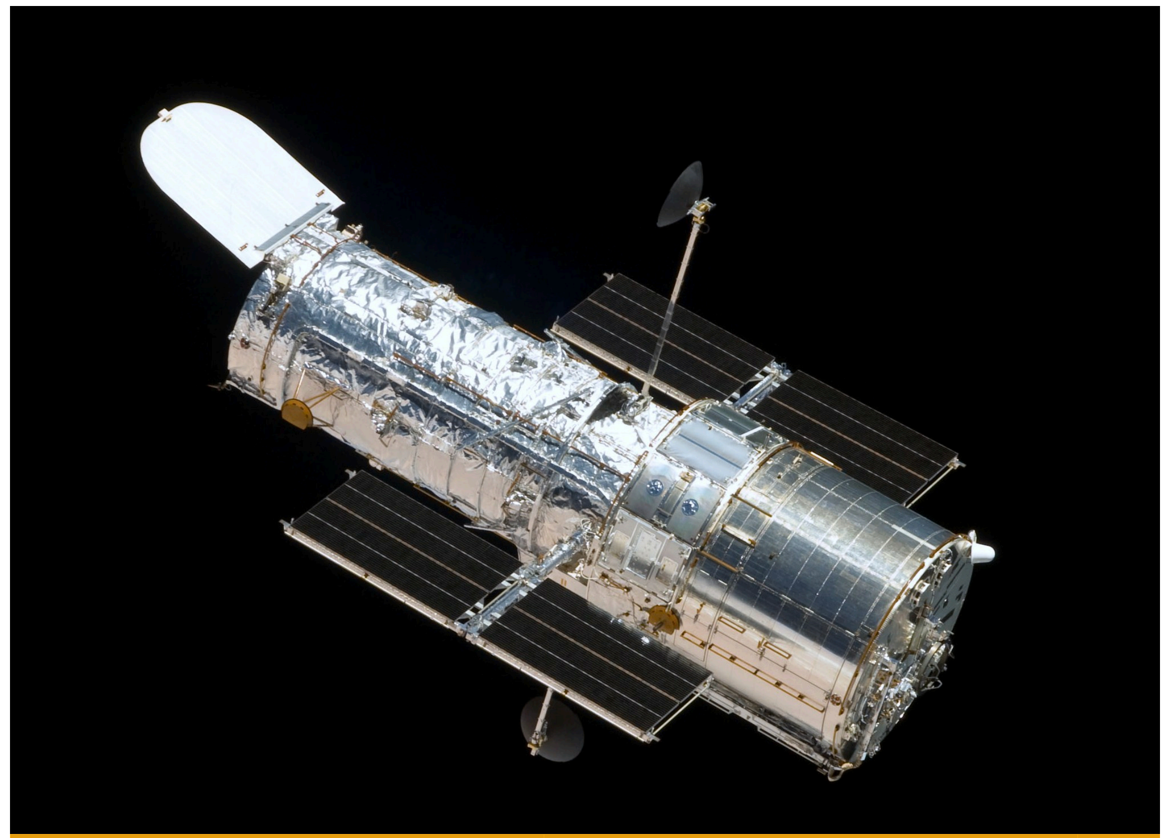
DRONES

What is a robot?

What is a robot?



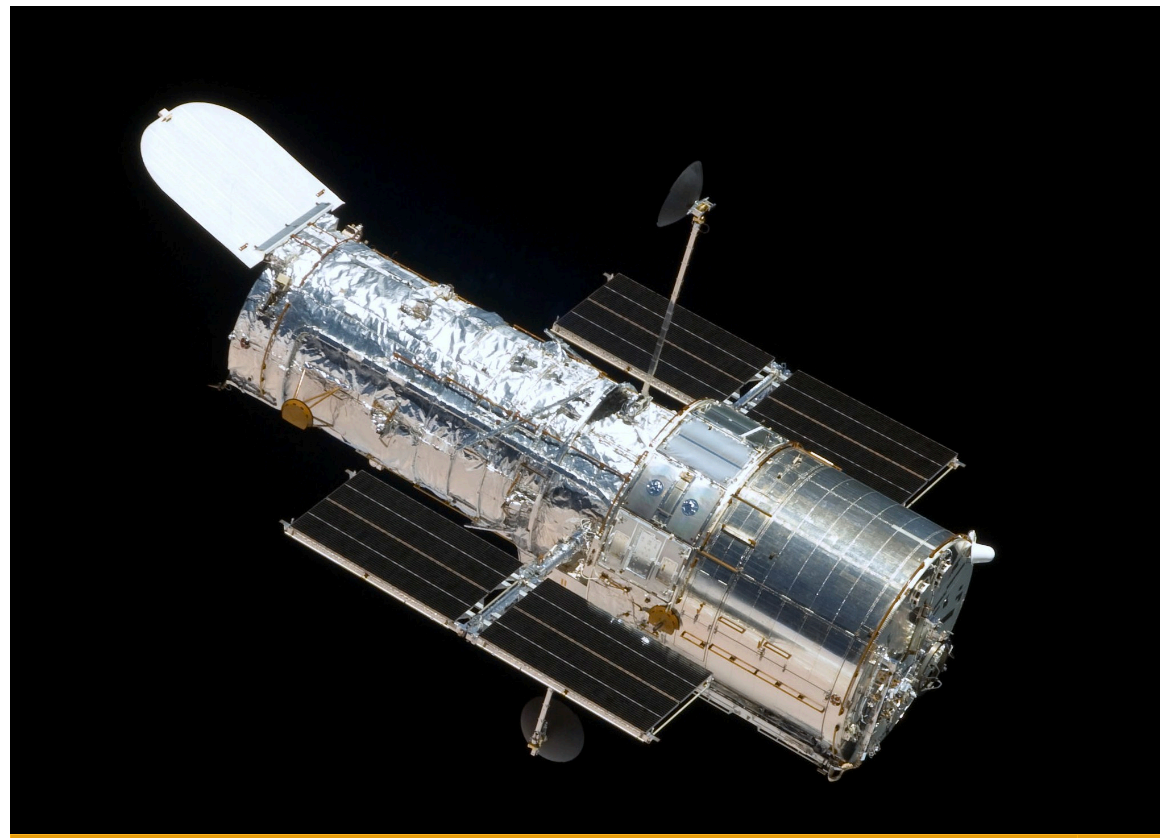
What is a robot?



TELESCOPES

MR IMAGERS

What is a robot?

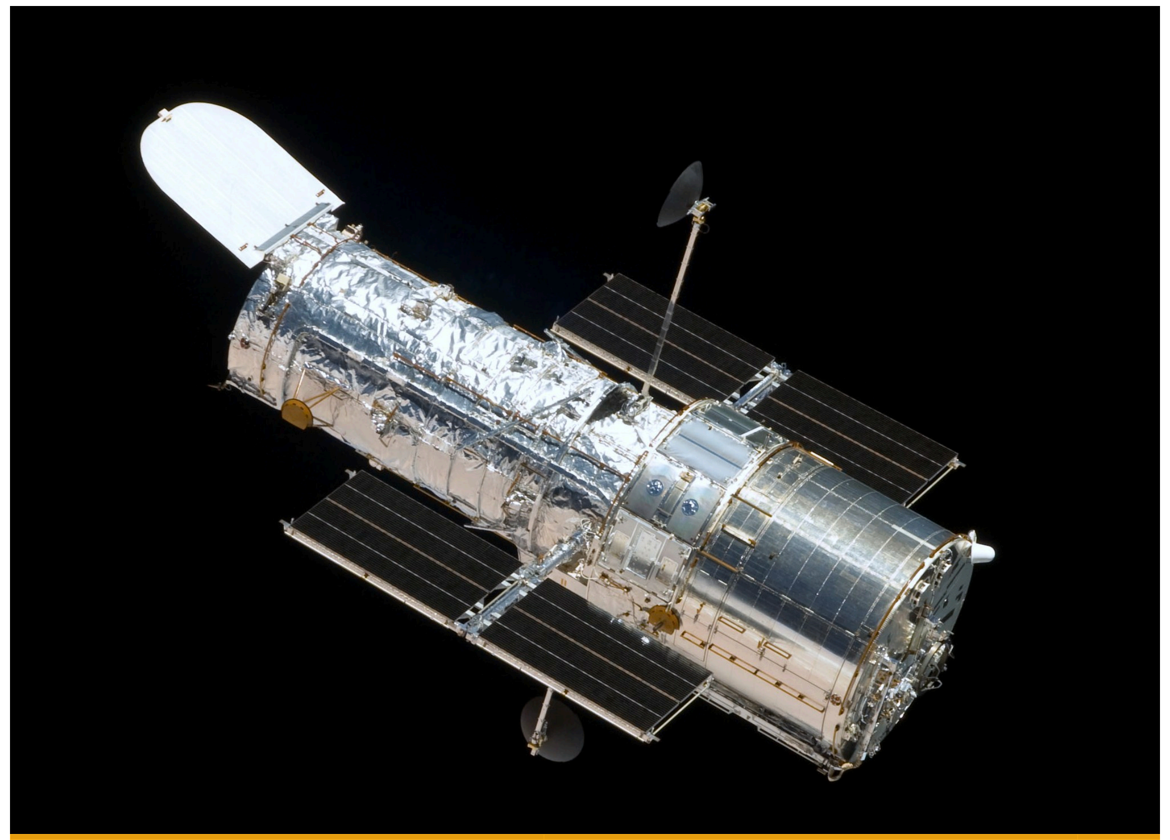


TELESCOPES

MR IMAGERS

SPECTROMETERS

What is a robot?

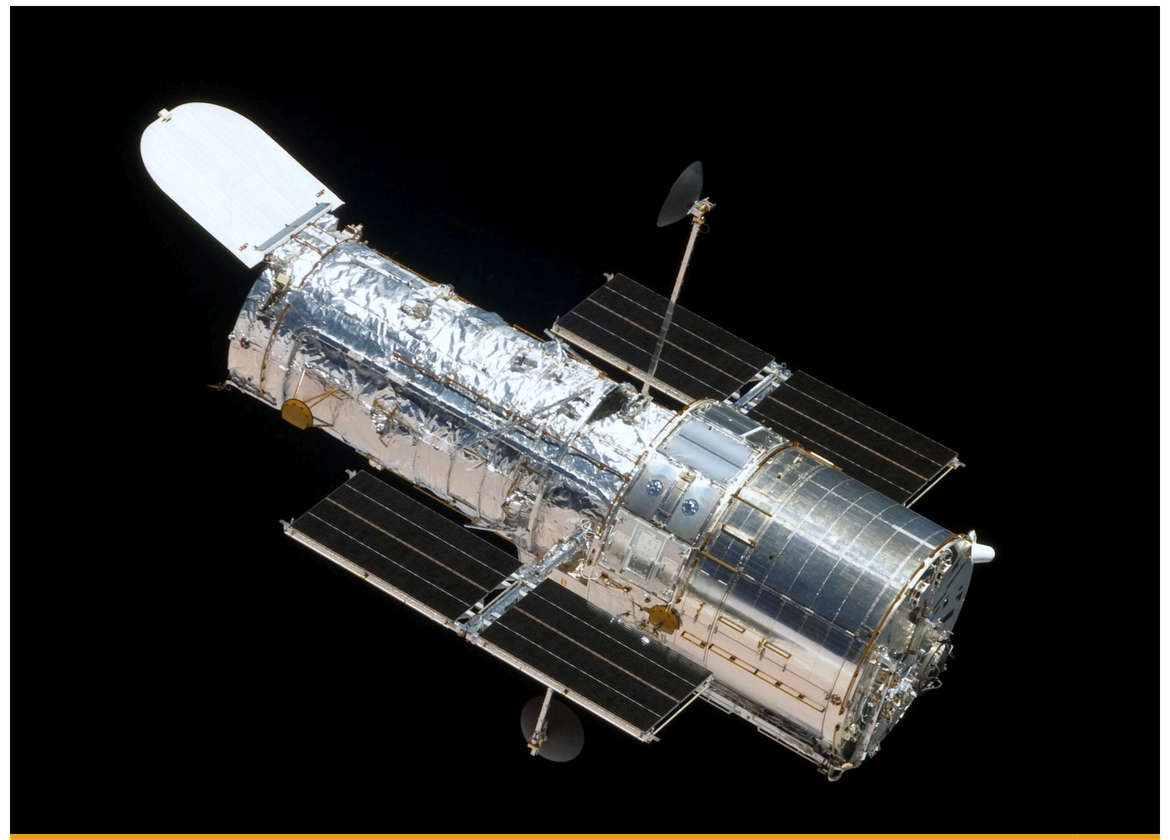

TELESCOPES


MR IMAGERS

SPECTROMETERS


CYTOMETERS

What is a robot?



TELESCOPES

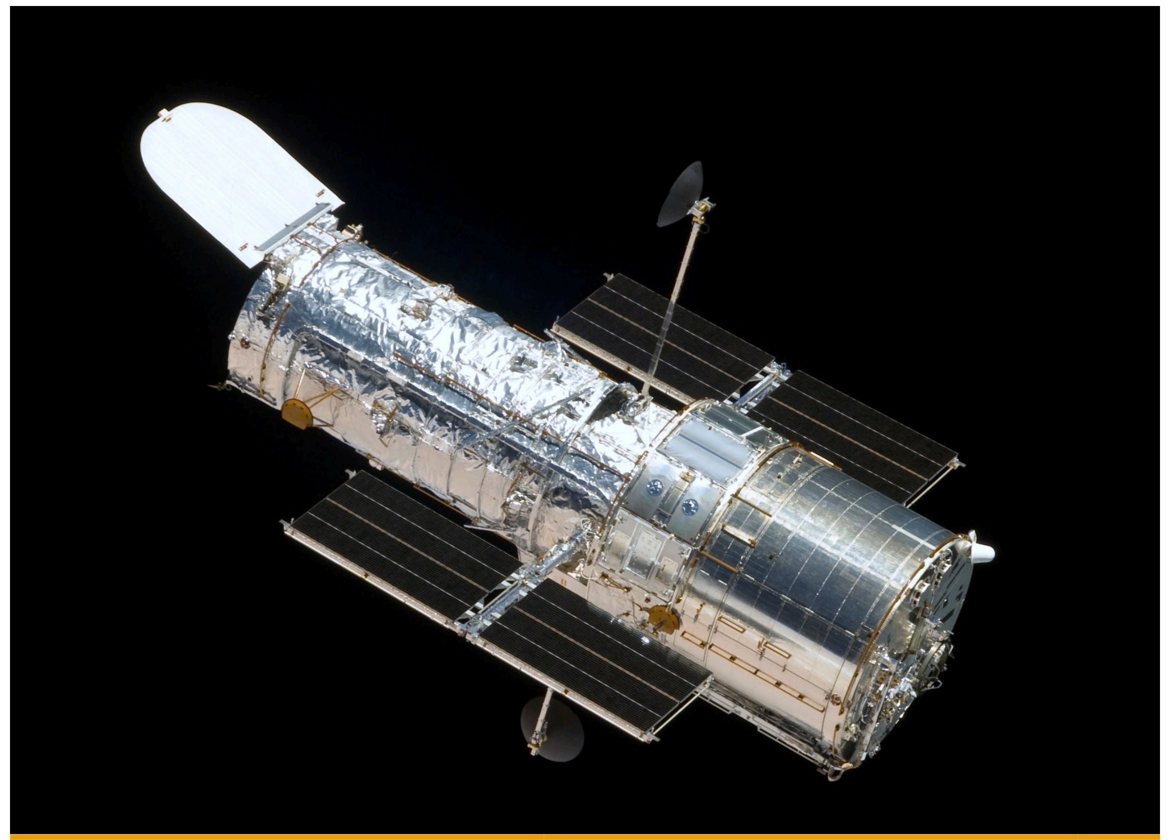
MR IMAGERS

SPECTROMETERS

CYTOMETERS

ACCELERATORS

What is a robot?



TELESCOPES

MR IMAGERS

SPECTROMETERS

CYTOMETERS

ACCELERATORS

...AND MORE

The future of AI + Measurement

Every scientific instrument is a robot and can be smarter

Tremendous opportunities for collaboration across U. Chicago and National Labs

Getting people in the same room is just a start — they have to speak each other's language!

ACCELERATORS

Argonne NATIONAL LABORATORY Fermilab THE UNIVERSITY OF CHICAGO

Extra Slides

Lack of automated analysis inhibits scientific advancement

Contemporary computational approaches for spectral analysis are effectively **library lookups**
(thus can only find “known knowns”)

Yet there are >**10^{60} possible** small molecules!
(obviously impossible to build comprehensive libraries)

Inhibits scale

80% of human small metabolites are still **unknown** [1]
(even worse for other organisms)

Crude oil can have **over 1,000,000 unique compounds** [2] and its composition is **still a mystery** [3]

Inhibits Robotic Laboratories

Automated synthetic chemistry and wet lab platforms are coming online

But how do we know what they made? Can’t close the loop if you can’t measure the output

Inhibits new instrumentation

Quantum Sensing and other molecule sensing techniques are increasingly viable

Often have fundamentally different tradeoffs from classical instrumentation, resulting in tremendous data interpretation challenges

[1] Dias, D., Jones, et. al. (2016). Current and Future Perspectives on the Structural Identification of Small Molecules in Biological Systems. Metabolites, 6(4), 46.

[2] Beens, J., Blomberg, J., & Schoenmakers, P. J. (2000). Proper Tuning of Comprehensive Two-Dimensional Gas Chromatography (GC×GC) to Optimize the Separation of Complex Oil Fractions. *Journal of High Resolution Chromatography*, *23*(3), 182–188.

[3] Panda, S. K., Andersson, J. T., & Schrader, W. (2009). Characterization of supercomplex crude oil mixtures: What is really in there? Angewandte Chemie - International Edition, 48(10), 1788–1791.

Key insight:

Spectroscopy is an inverse problem

What is an inverse problem?

What is an inverse problem?

Model System

Physical properties,
unknowns

Observables

Measurements
and data

What is an inverse problem?

The **forward** problem

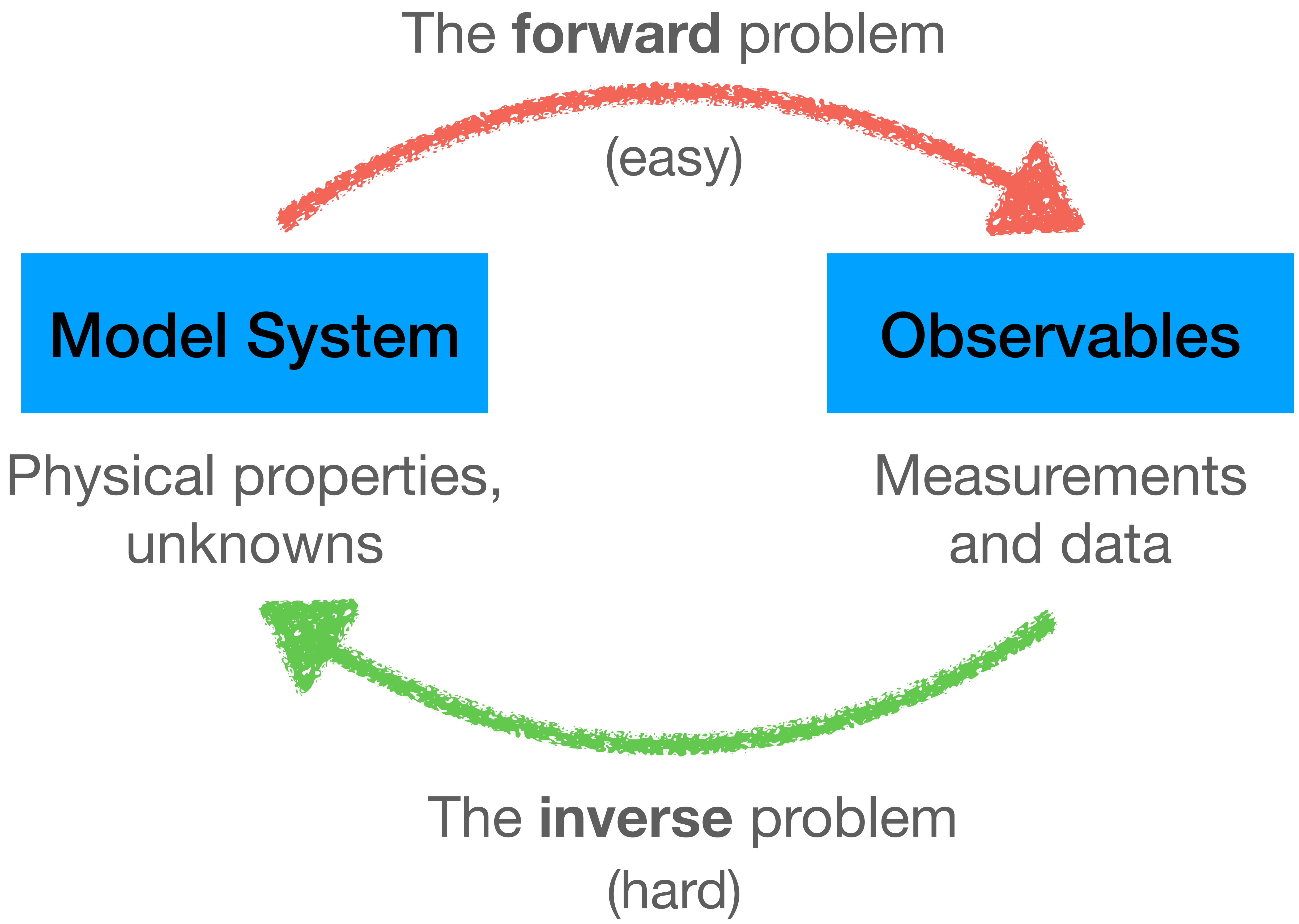
(easy)

Model System → **Observables**

Physical properties,
unknowns

Measurements
and data

What is an inverse problem?



What is an inverse problem?

The **forward** problem

(easy)

Model System

Physical properties, unknowns

Observables

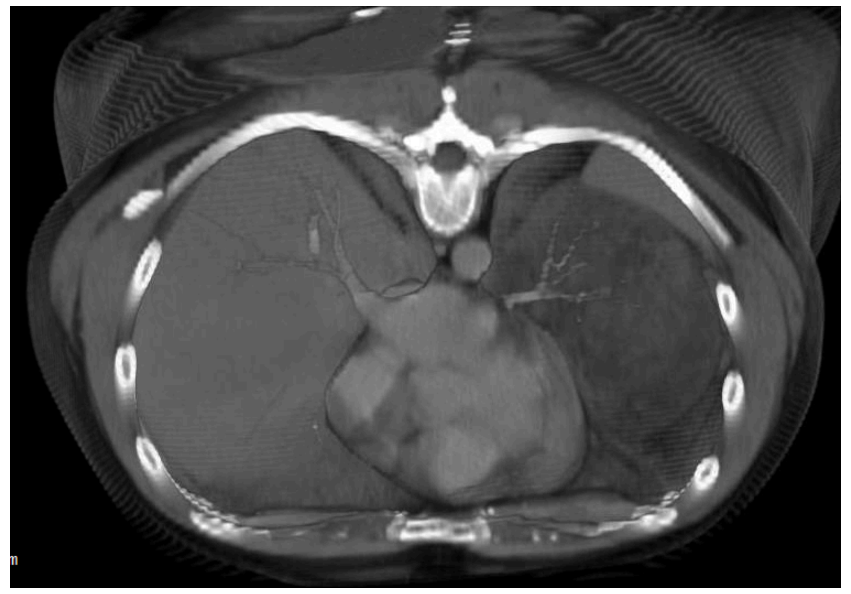
Measurements and data

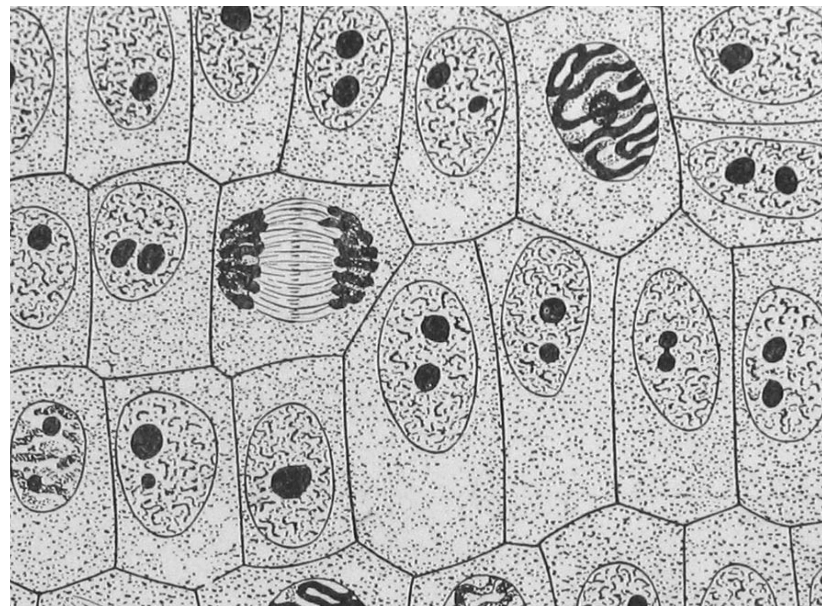
The **inverse** problem

(hard)

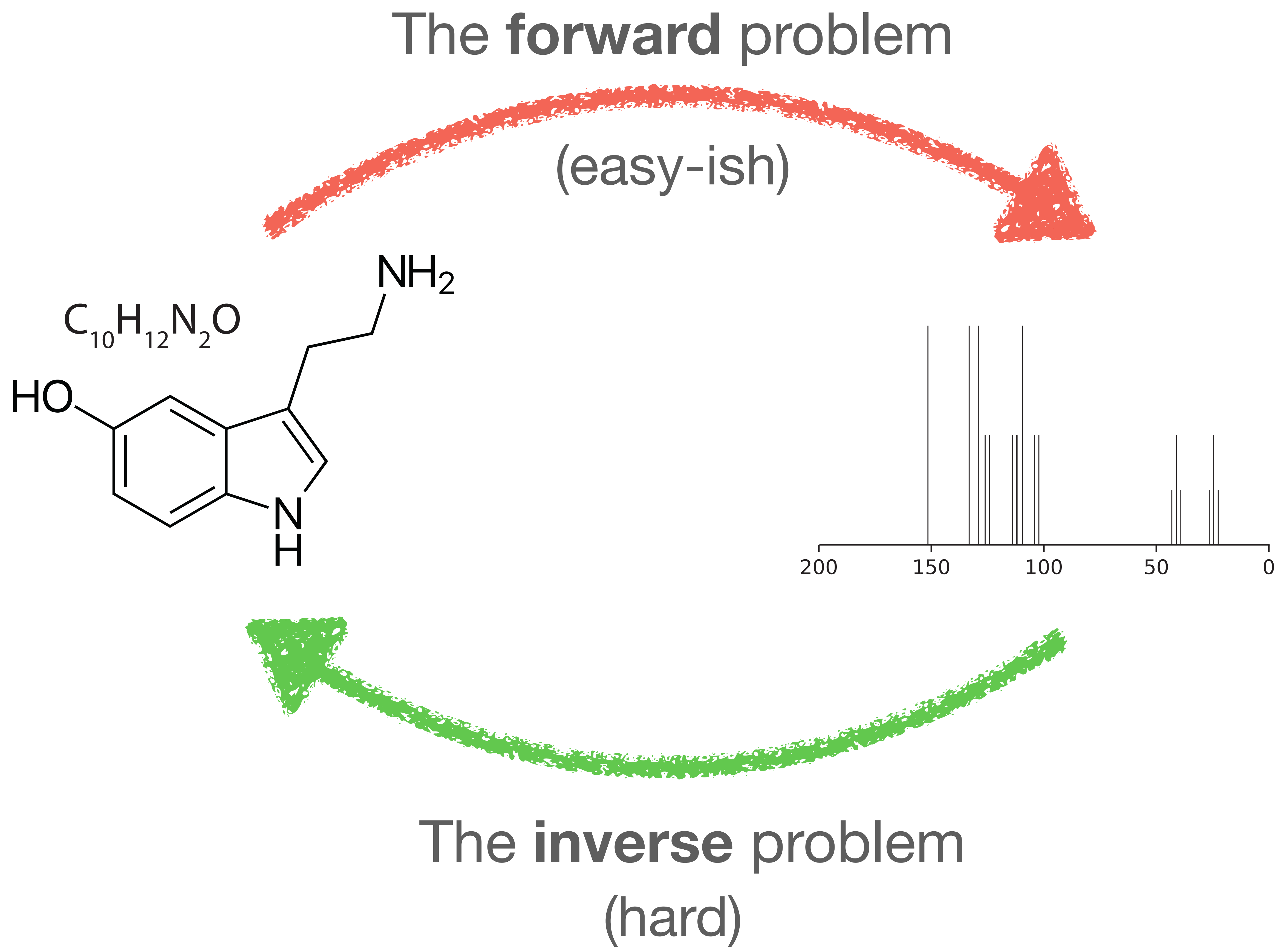
20th Century Measurement

Linear, continuous inverse problems transformed measurement in the latter half of the 20th century

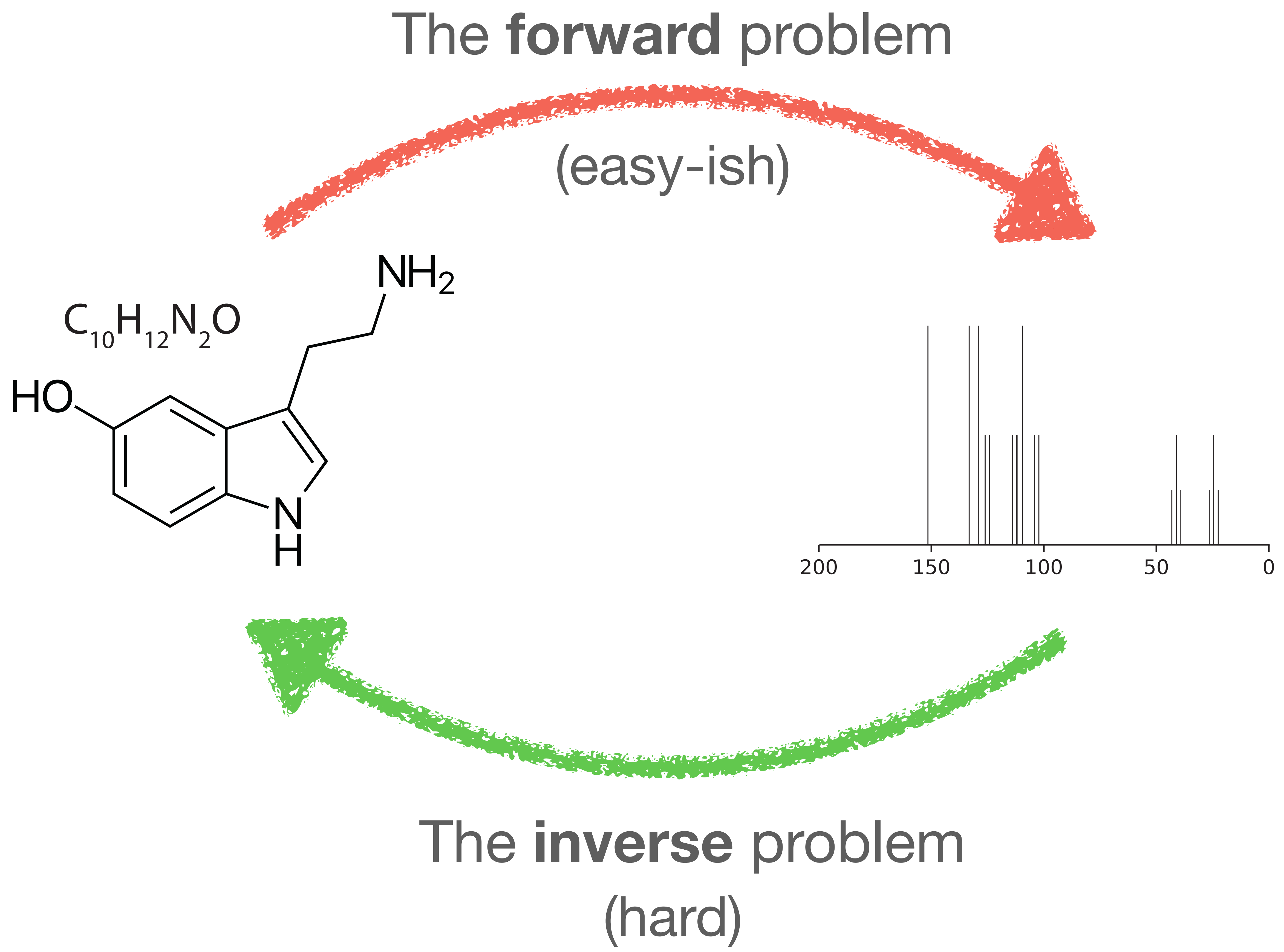




Molecular spectroscopy as an inverse problem



Molecular spectroscopy as an inverse problem



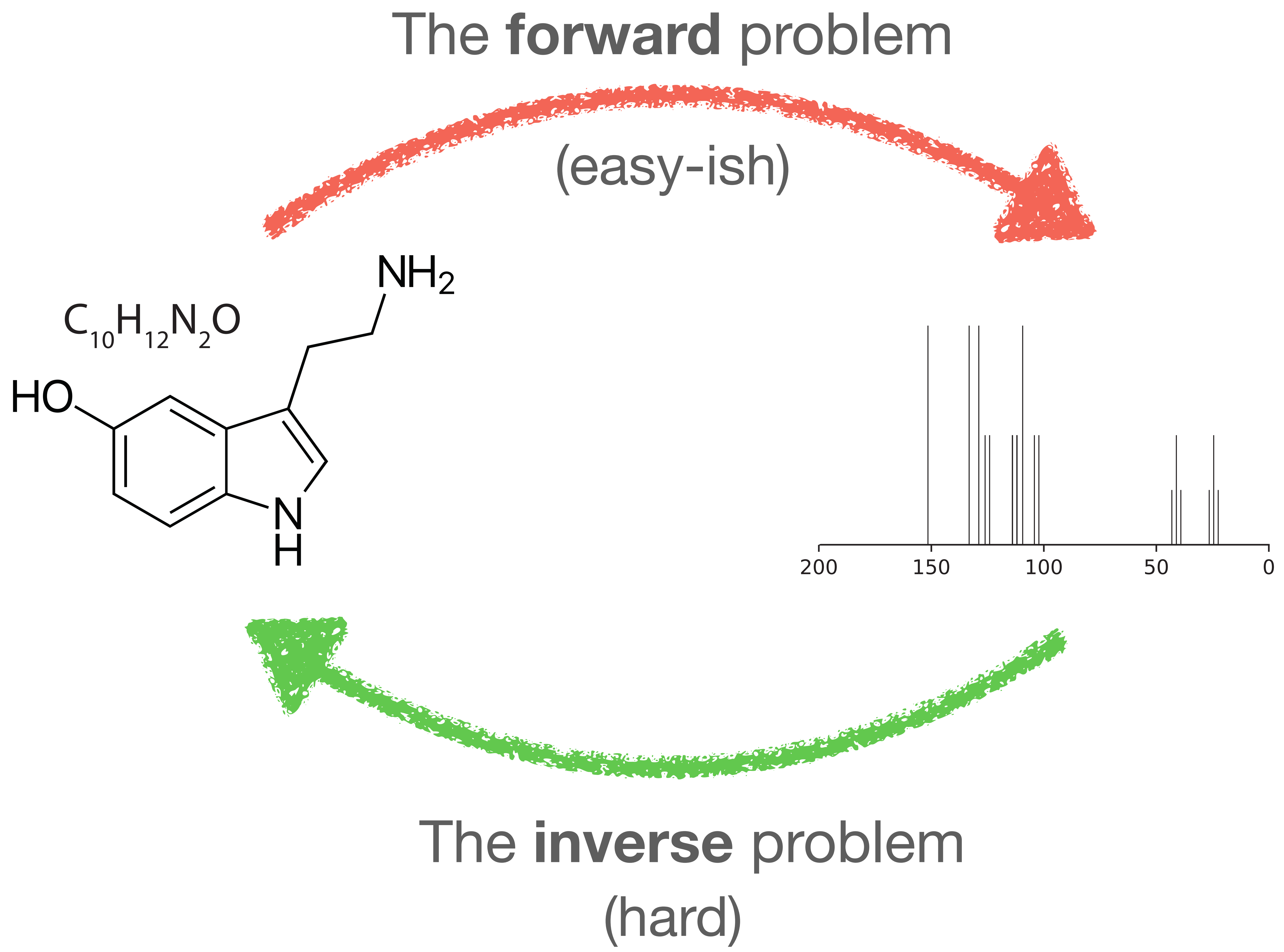
Forward

Calculating the spectrum for a given structure

SCF/DFT solves the forward problem for many modalities of interest

Calling this “easy” is a stretch — performance is cubic in the number of atoms and many aspects of experimental setup (conformational diversity, salvation, etc.) are still challenging.

Molecular spectroscopy as an inverse problem



Forward

Calculating the spectrum for a given structure

SCF/DFT solves the forward problem for many modalities of interest

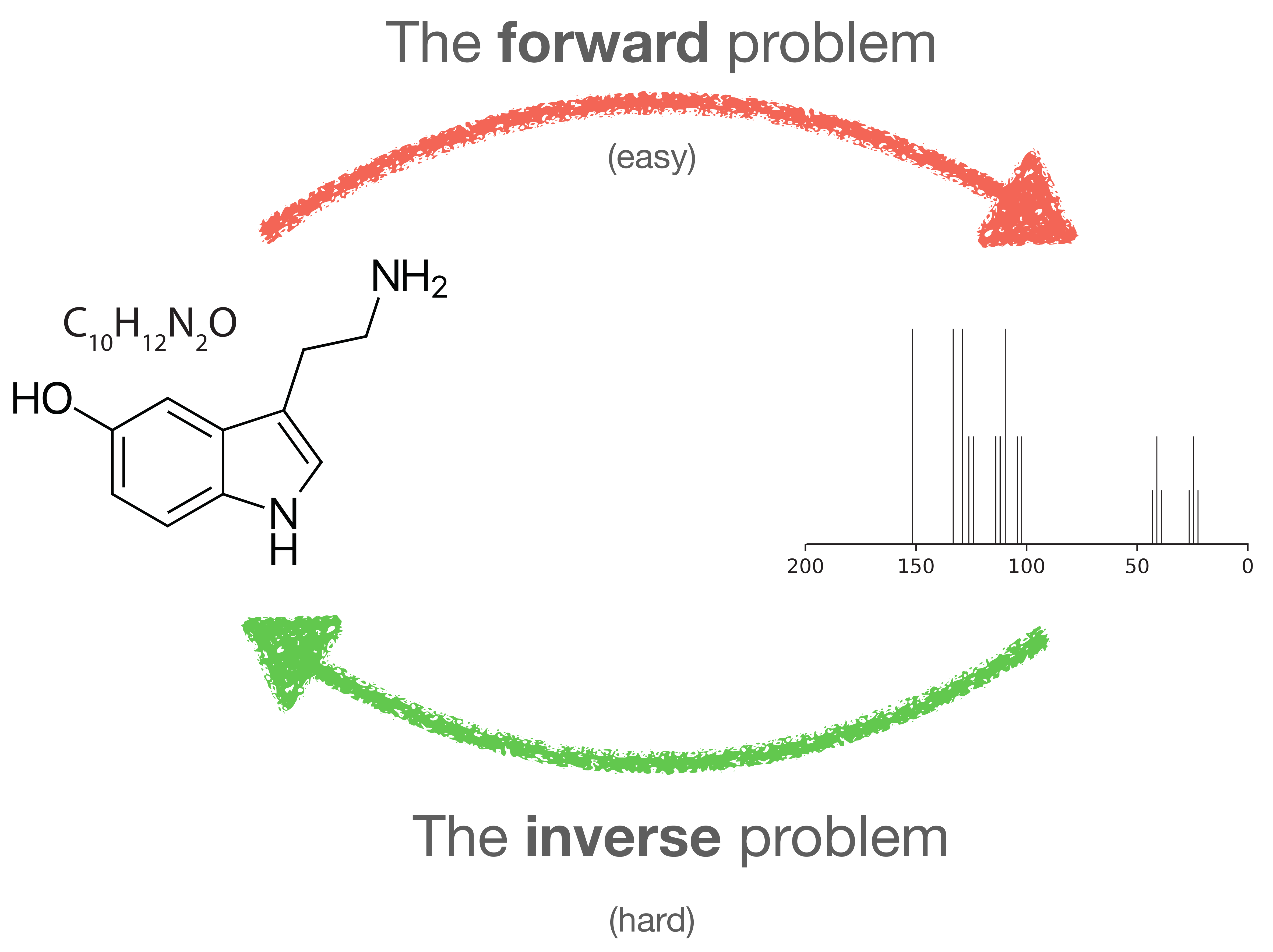
Calling this "easy" is a stretch — performance is cubic in the number of atoms and many aspects of experimental setup (conformational diversity, salvation, etc.) are still challenging.

Inverse

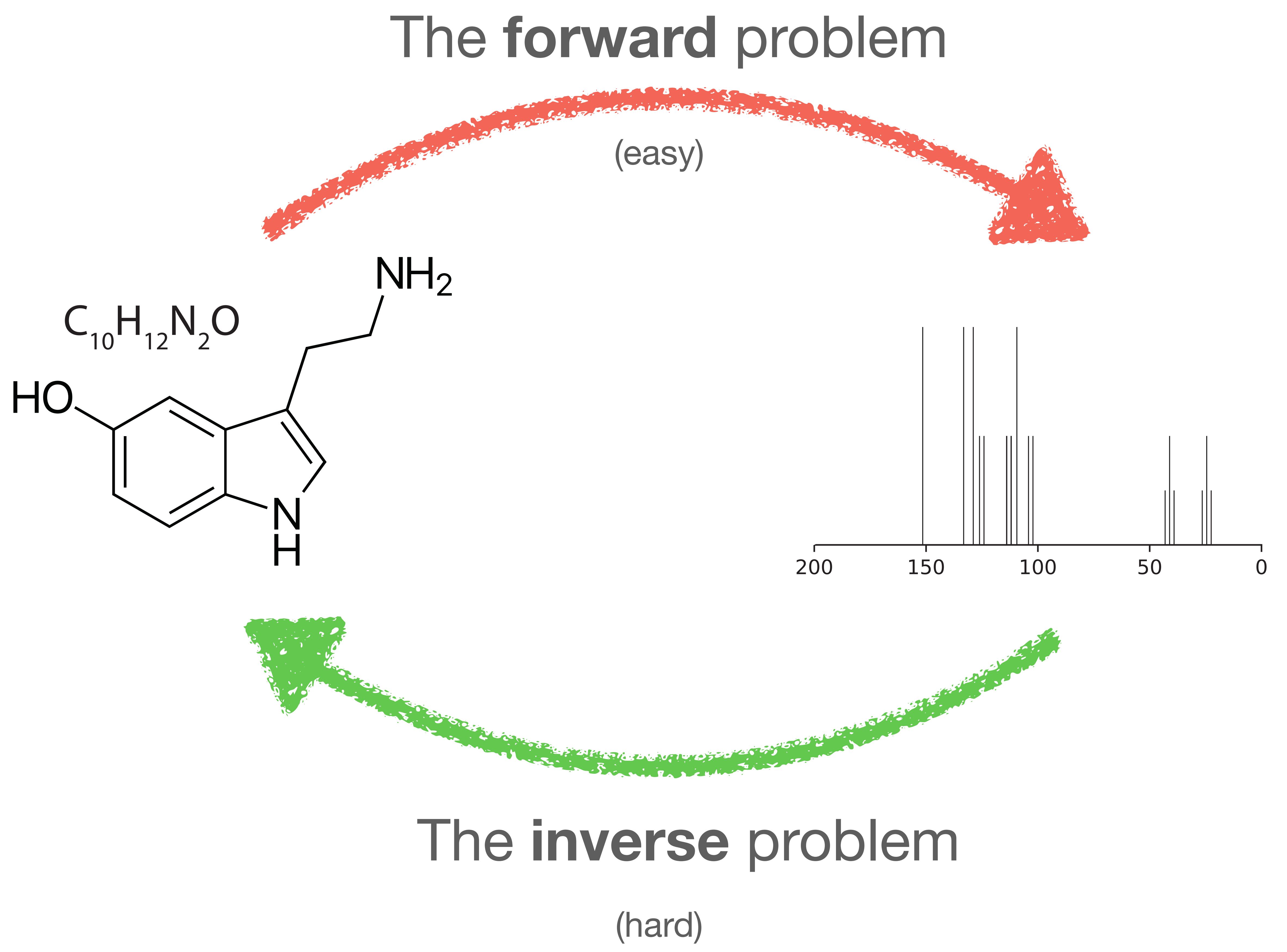
Deducing the structure for a given spectrum

This is incredibly challenging, a long-standing open problem

Highly nonlinear forward model
Combinatorial solution space
Single correct structure!



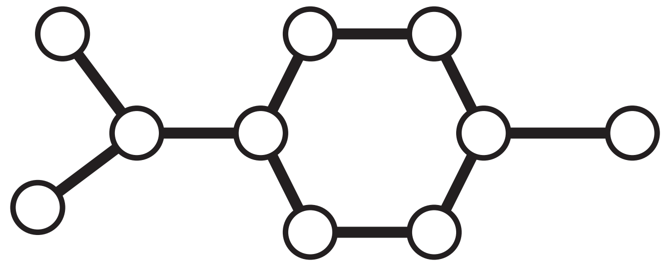
The **forward** problem
(easy)
$C_{10}H_{12}N_2O$
NH_2
HO
N
H
200
150
100
50
0
The **inverse** problem
(hard)

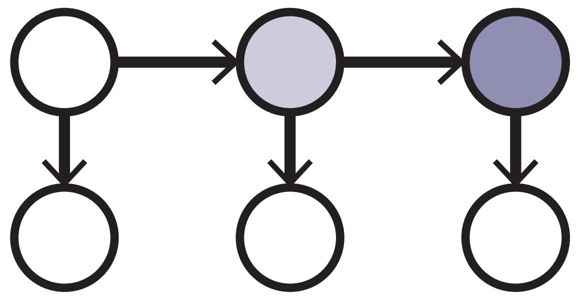


AI advances make this possible

We’re developing AI techniques to solve this.

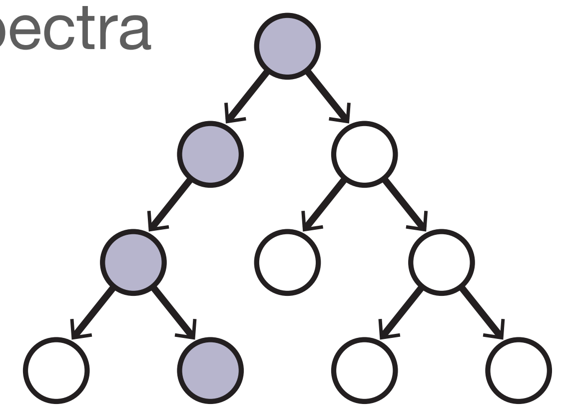
Physics-informed deep learning and graph neural networks let us generate millions of synthetic spectra for training data

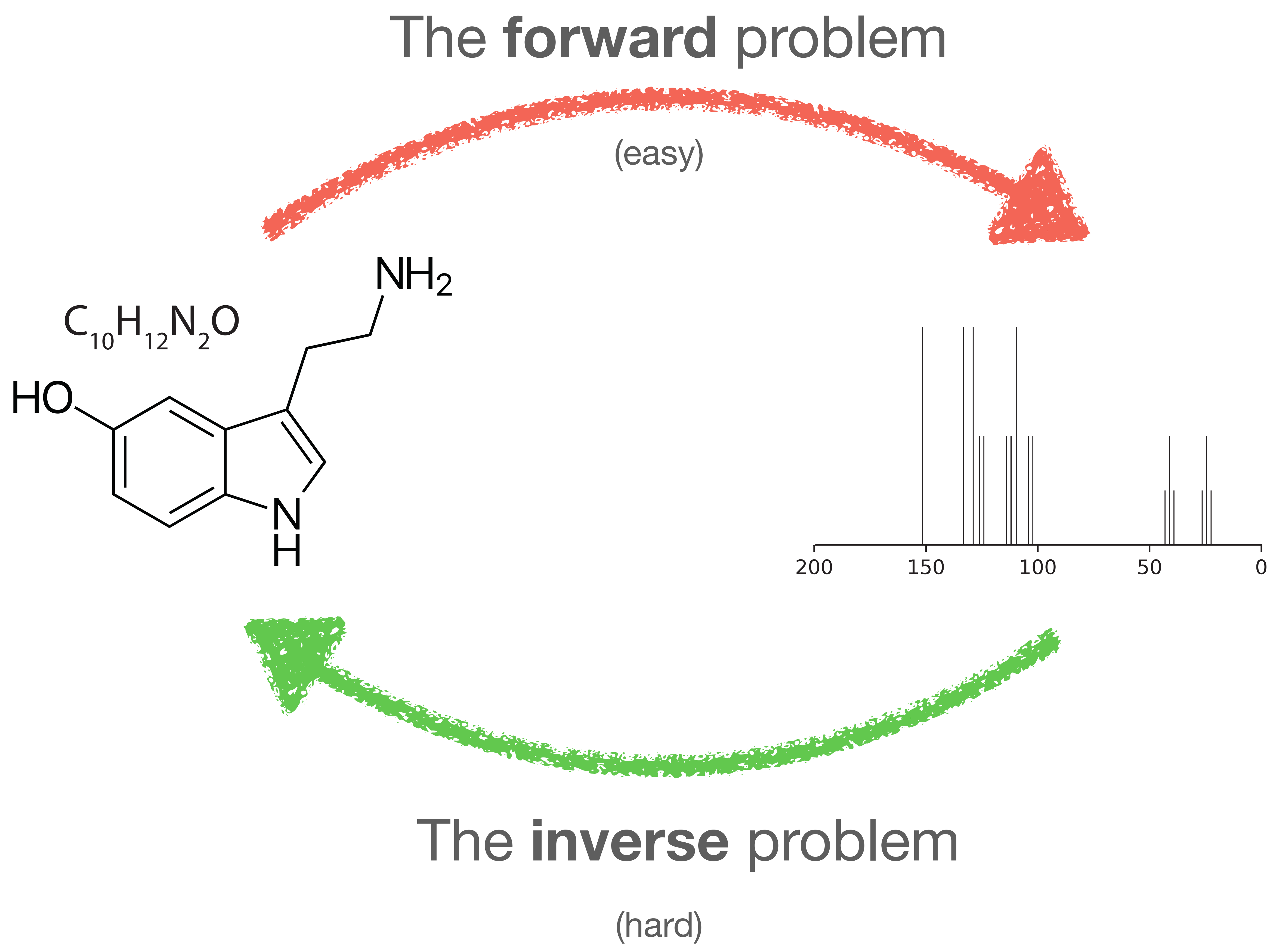




Structured prediction via deep imitation learning lets us learn to build molecules consistent with observed spectra

Deep latent variable models let us model *and understand* physical measurement processes where ab initio techniques fail.

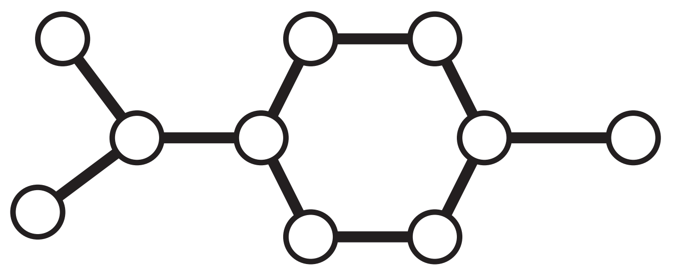


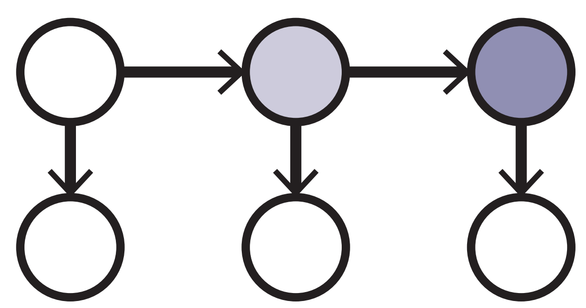


AI advances make this possible

We're developing AI techniques to solve this.

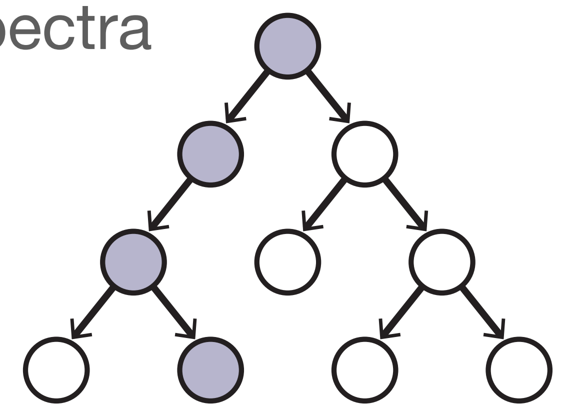
Physics-informed deep learning and graph neural networks let us generate millions of synthetic spectra for training data





Structured prediction via deep imitation learning lets us learn to build molecules consistent with observed spectra

Deep latent variable models let us model *and understand* physical measurement processes where ab initio techniques fail.



Research Objectives

We've had early success in NMR and are moving into mass spec

	NMR	MS
Forward	Fast model to simulate spectra: DFT accuracy in milliseconds	Computational forward model still an open research challenge!
Inverse	Can predict correct structure with high accuracy on a wide variety of compounds from 1D 13C spectrum	Next big challenge

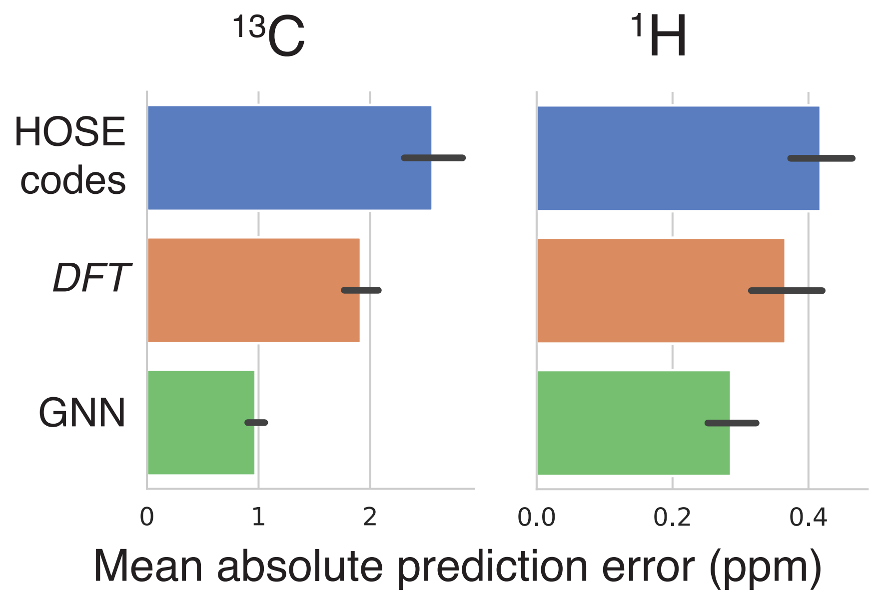
Methodologies

We learn to build molecules from spectroscopic data by first taking them apart

Methodologies

We learn to build molecules from spectroscopic data by first taking them apart

1. We need a lot of training data!



Solution: build a fast approximation to the forward model that lets you generate 100 M synthetic spectra

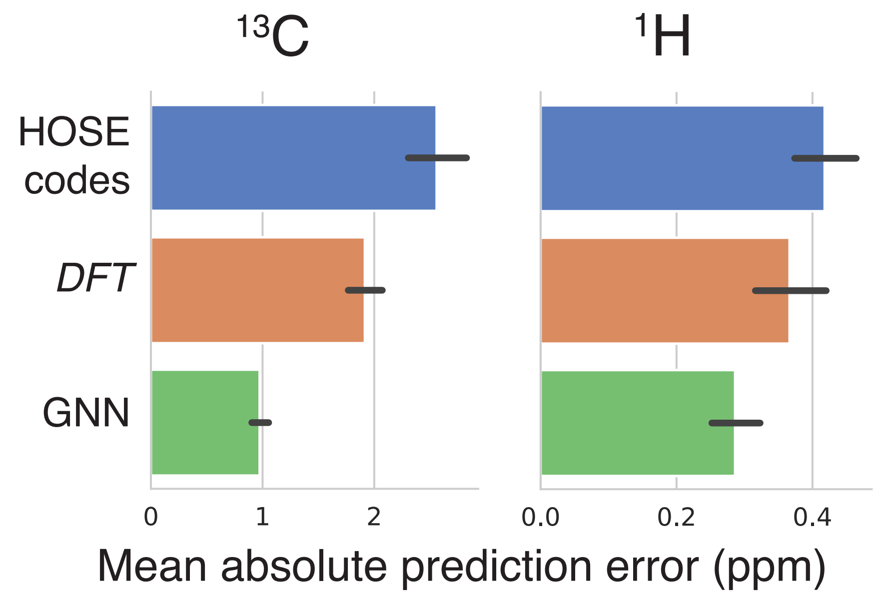
(Bootstrapped from 30k experimental spectra)

[1] Jonas, Kuhn. Rapid prediction of NMR spectral properties with quantified uncertainty. Journal of Cheminformatics, 11(1): 2019.

Methodologies

We learn to build molecules from spectroscopic data by first taking them apart

1. We need a lot of training data!



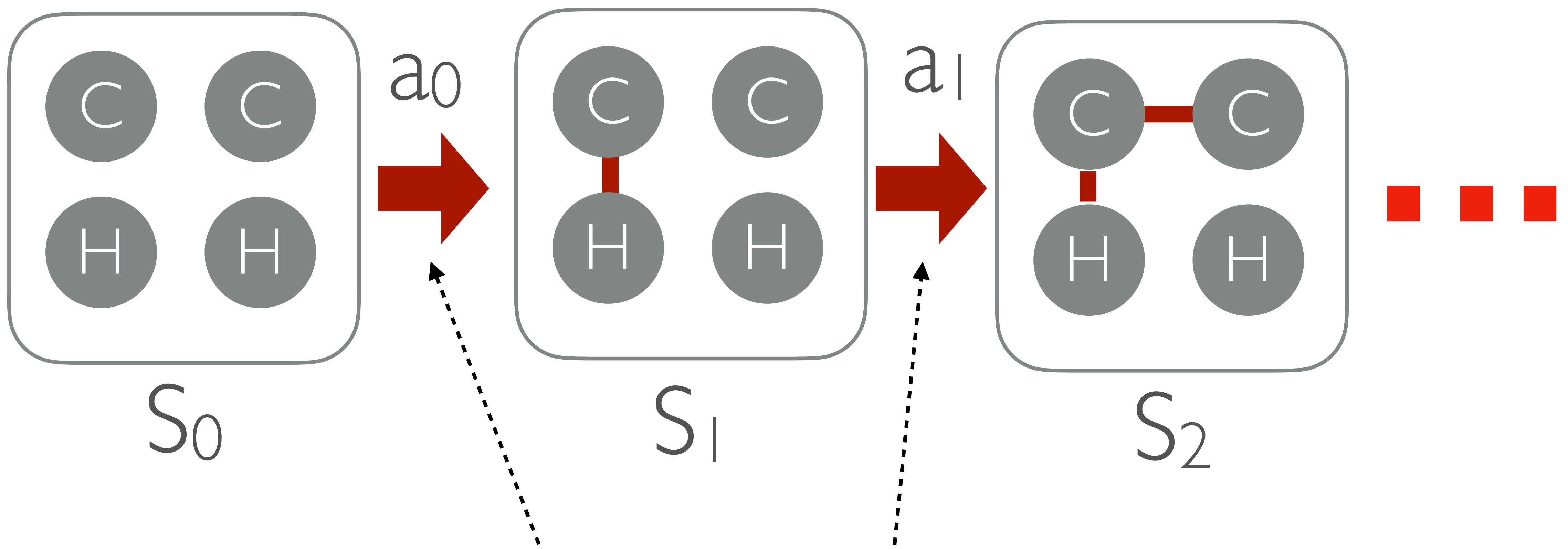
Solution: build a fast approximation to the forward model that lets you generate 100 M synthetic spectra

(Bootstrapped from 30k experimental spectra)

[1] Jonas, Kuhn. Rapid prediction of NMR spectral properties with quantified uncertainty. Journal of Cheminformatics, 11(1): 2019.

2. Generating complete molecules is hard!

Solution: Construct molecule incrementally — use deep imitation learning to learn to place the next bond of a partial molecule [2]



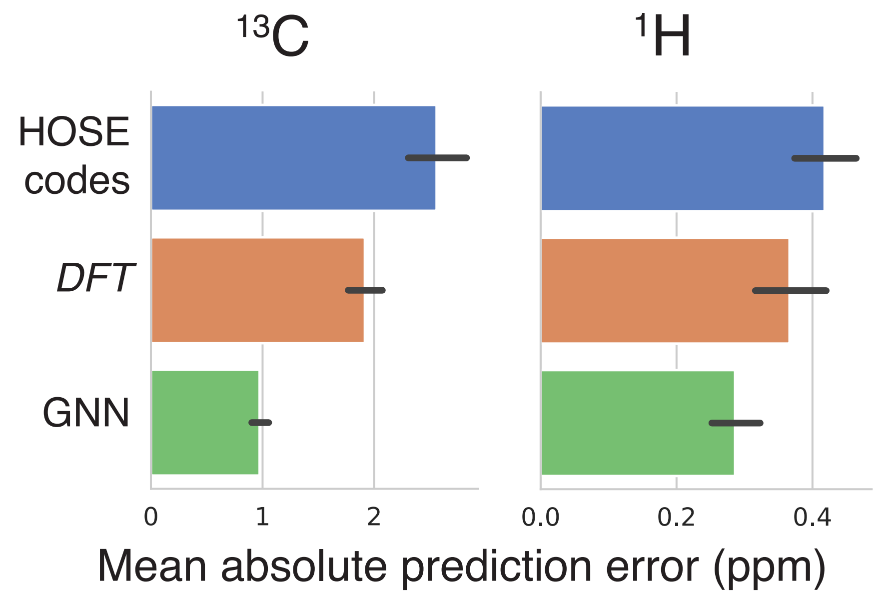
Learn this function: $p(a_k \mid S_k, \text{spectrum})$

[2] Jonas, Deep Imitation learning for Molecular Inverse Problems, NeurIPS 2019

Methodologies

We learn to build molecules from spectroscopic data by first taking them apart

1. We need a lot of training data!



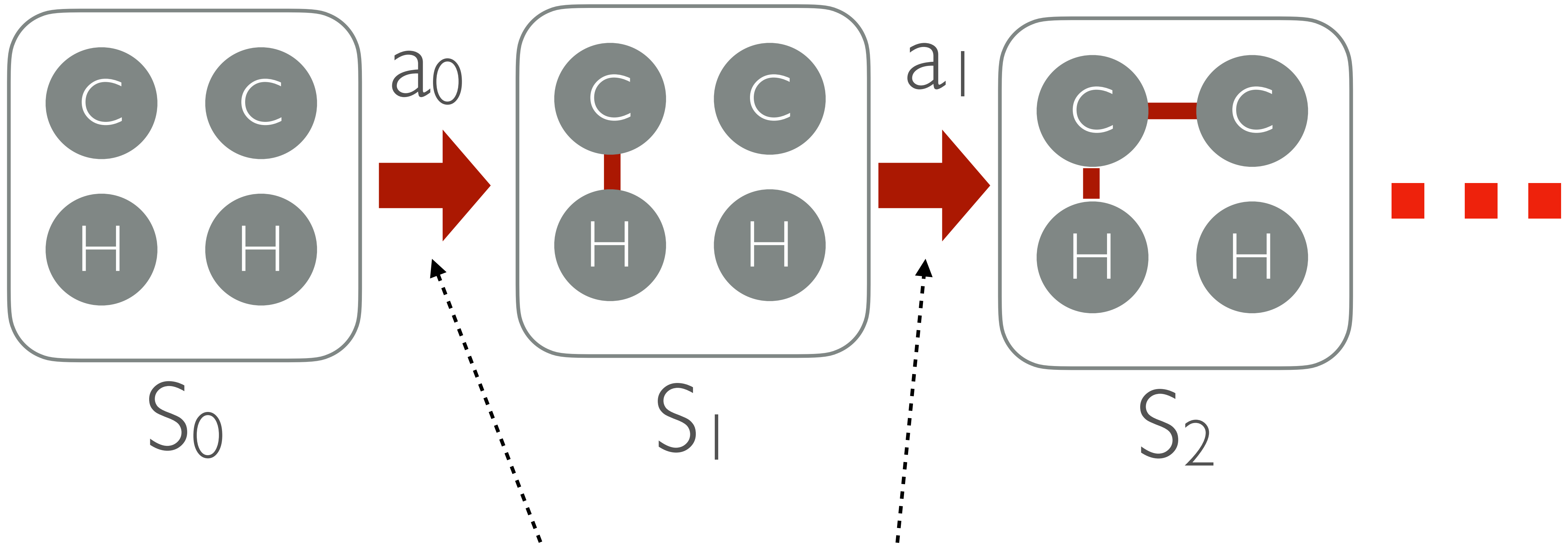
Solution: build a fast approximation to the forward model that lets you generate 100 M synthetic spectra

(Bootstrapped from 30k experimental spectra)

[1] Jonas, Kuhn. Rapid prediction of NMR spectral properties with quantified uncertainty. Journal of Cheminformatics, 11(1): 2019.

2. Generating complete molecules is hard!

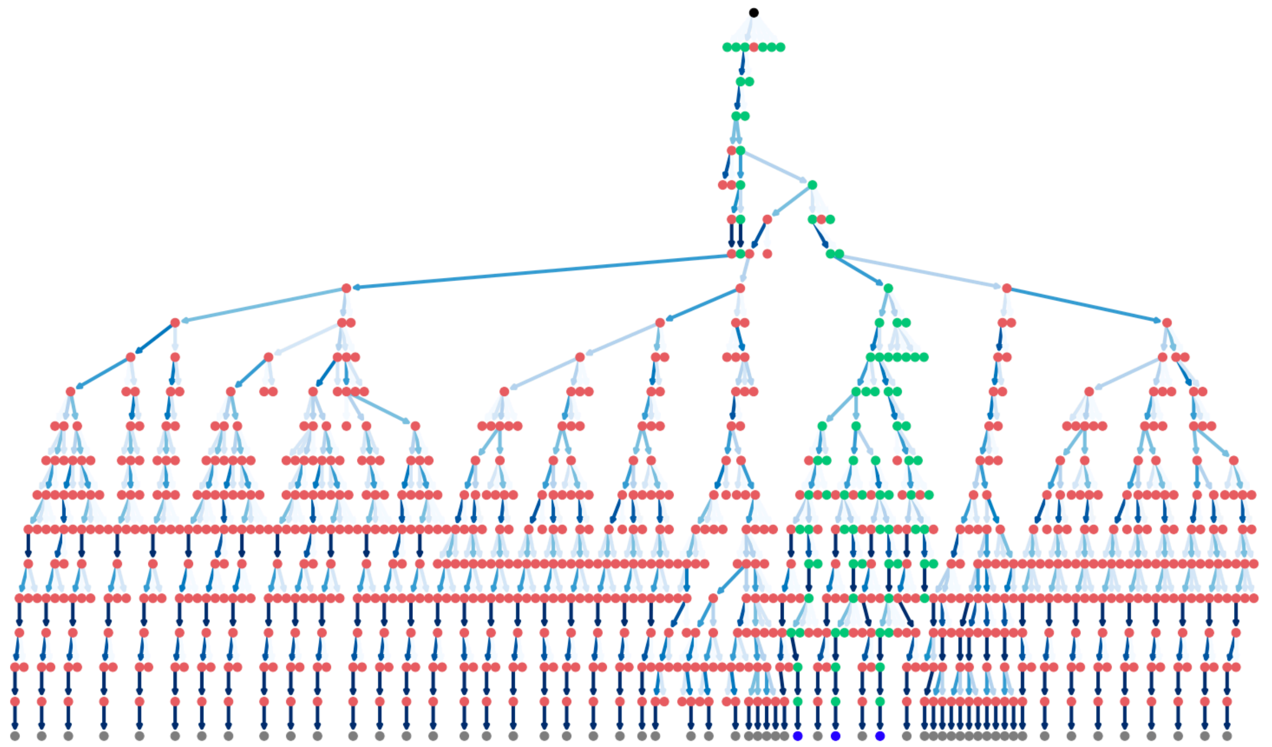
Solution: Construct molecule incrementally — use deep imitation learning to learn to place the next bond of a partial molecule [2]



Learn this function: $p(a_k \mid S_k, \text{spectrum})$

[2] Jonas, Deep Imitation learning for Molecular Inverse Problems, NeurIPS 2019

3.Generate candidate structures

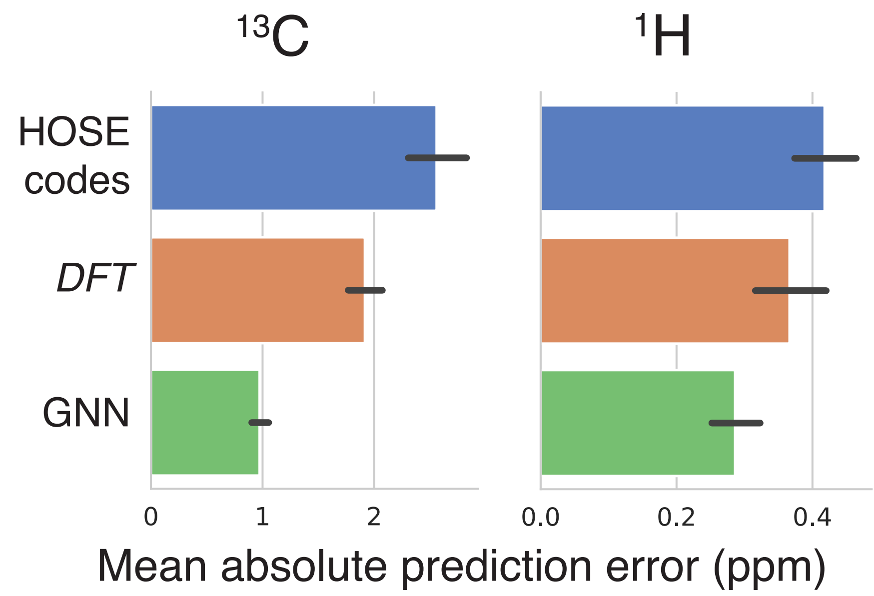


Search tree generating candidate structures from observed spectrum using learned function

Methodologies

We learn to build molecules from spectroscopic data by first taking them apart

1. We need a lot of training data!



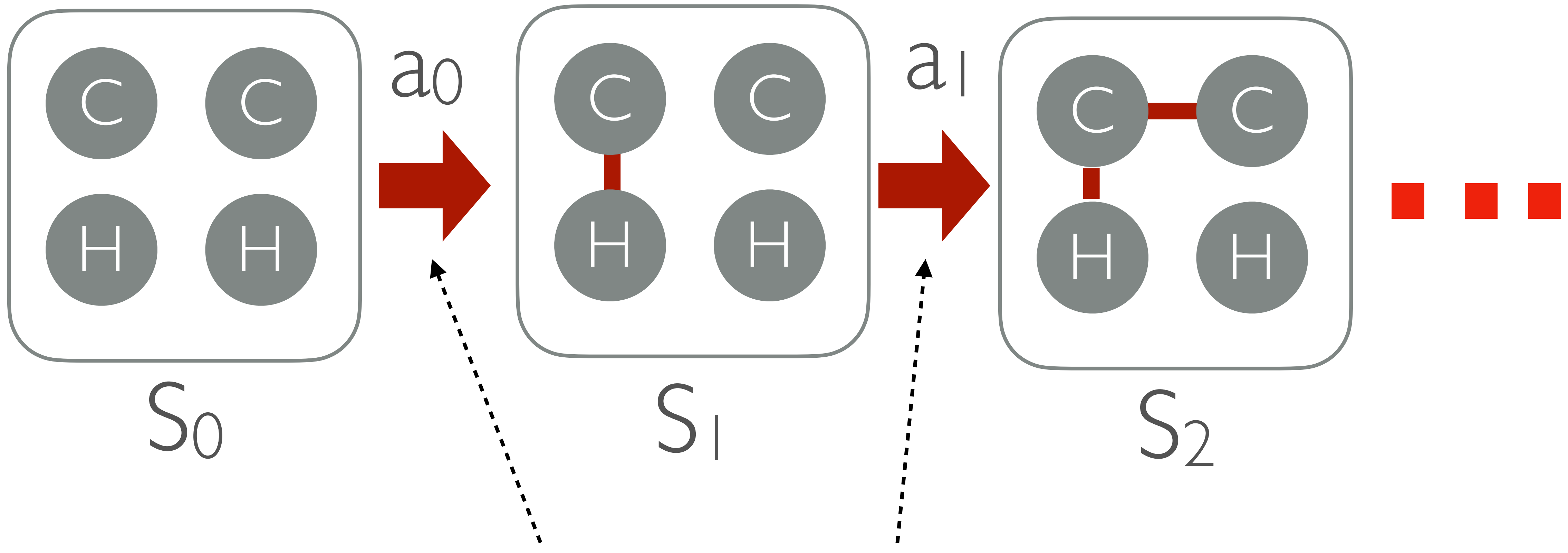
Solution: build a fast approximation to the forward model that lets you generate 100 M synthetic spectra

(Bootstrapped from 30k experimental spectra)

[1] Jonas, Kuhn. Rapid prediction of NMR spectral properties with quantified uncertainty. Journal of Cheminformatics, 11(1): 2019.

2. Generating complete molecules is hard!

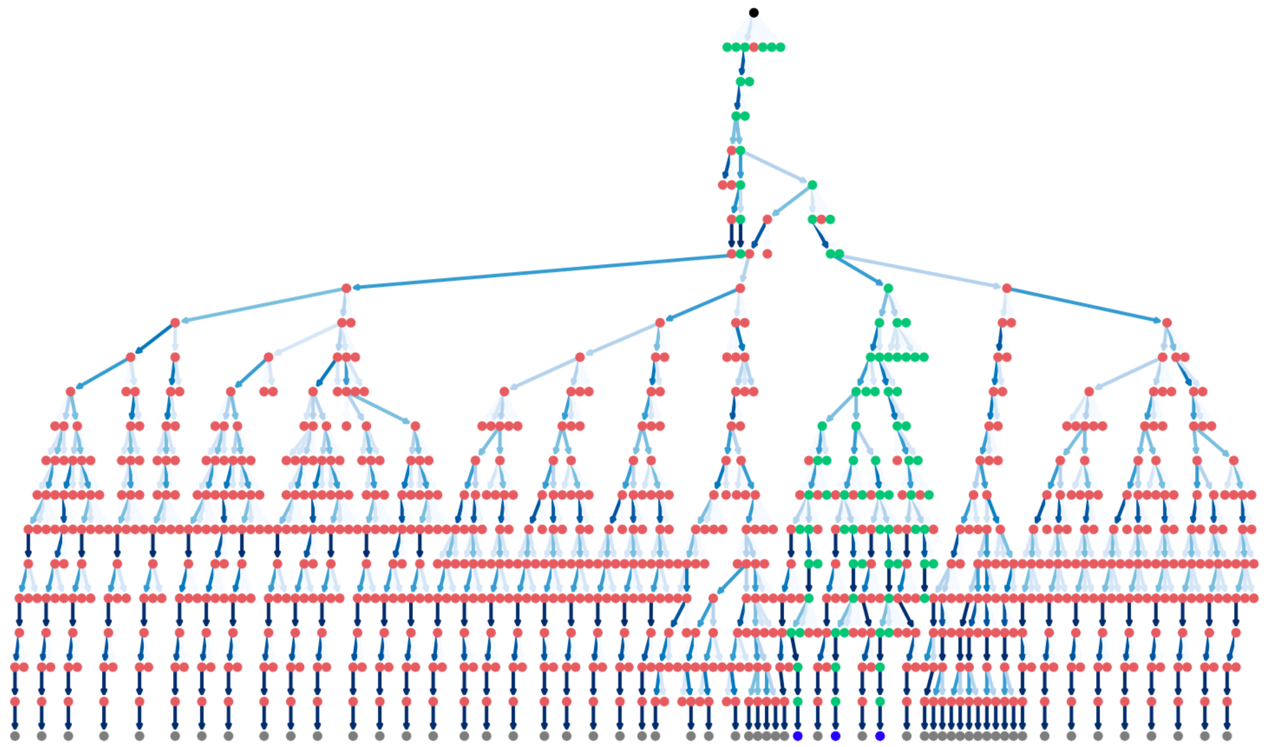
Solution: Construct molecule incrementally — use deep imitation learning to learn to place the next bond of a partial molecule [2]



Learn this function: $p(a_k \mid S_k, \text{spectrum})$

[2] Jonas, Deep Imitation learning for Molecular Inverse Problems, NeurIPS 2019

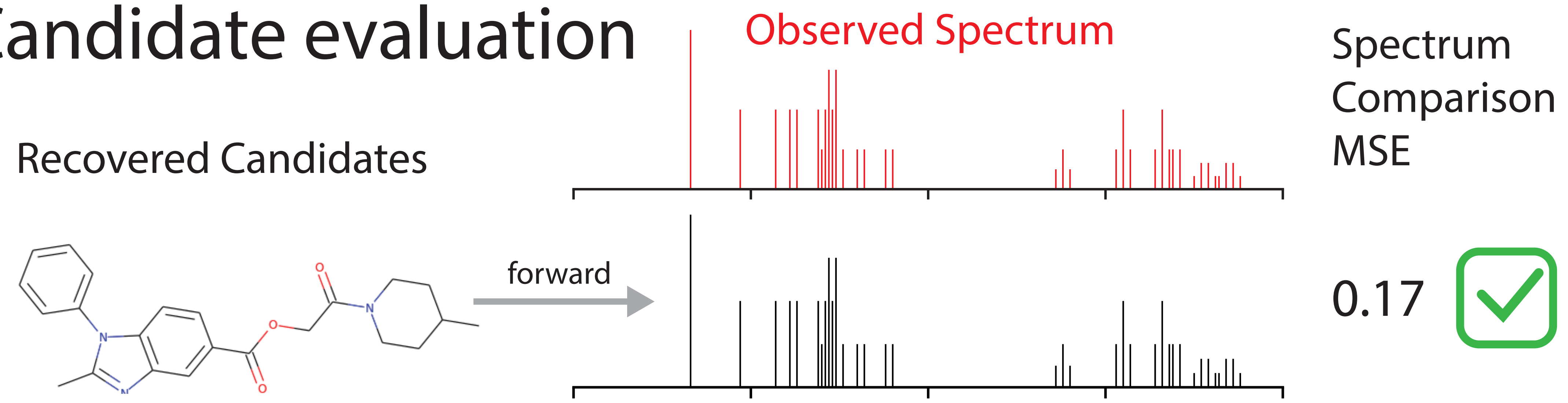
3.Generate candidate structures



Search tree generating candidate structures from observed spectrum using learned function

4. Use fast forward model to validate!

Candidate evaluation



We predict the right structure 68% of the time (up from 56% in 2020)

(96% of the time on most-confident mols!)

Three Phases

Build fast precise forward model

Real-time solution to Inverse problem

Compute optimal next measurement

Three Phases

Build fast precise forward model

Real-time solution to Inverse problem

Compute optimal next measurement

Why start with commodity spectroscopic modalities like NMR and MS?

Existing Data

NMRShiftDB : 50k NMR exp

NIST-17 : 250k GC/EI-MS exp

SDBS: 20k NMR exp

MassBank: 50k LC-MS/MS exp

Reliable Ubiquitous Hardware

At UChicago we have:

7 NMR specrometers (Bruker 400MHz+)

8 MS instruments (GC-EI/MS, QTOF, incoming Thermo Orbitraps)

Existing Platforms are programmable

Custom real-time pulse sequence design via Bruker hardware

MS HW enables programmatic control over collision energies and peak selection for fragmentation

Developing new techniques for acquisition

Modern AI requires tons of data

We are developing solutions to generate massive quantities of training data called

Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

We are developing solutions to generate massive quantities of training data called

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

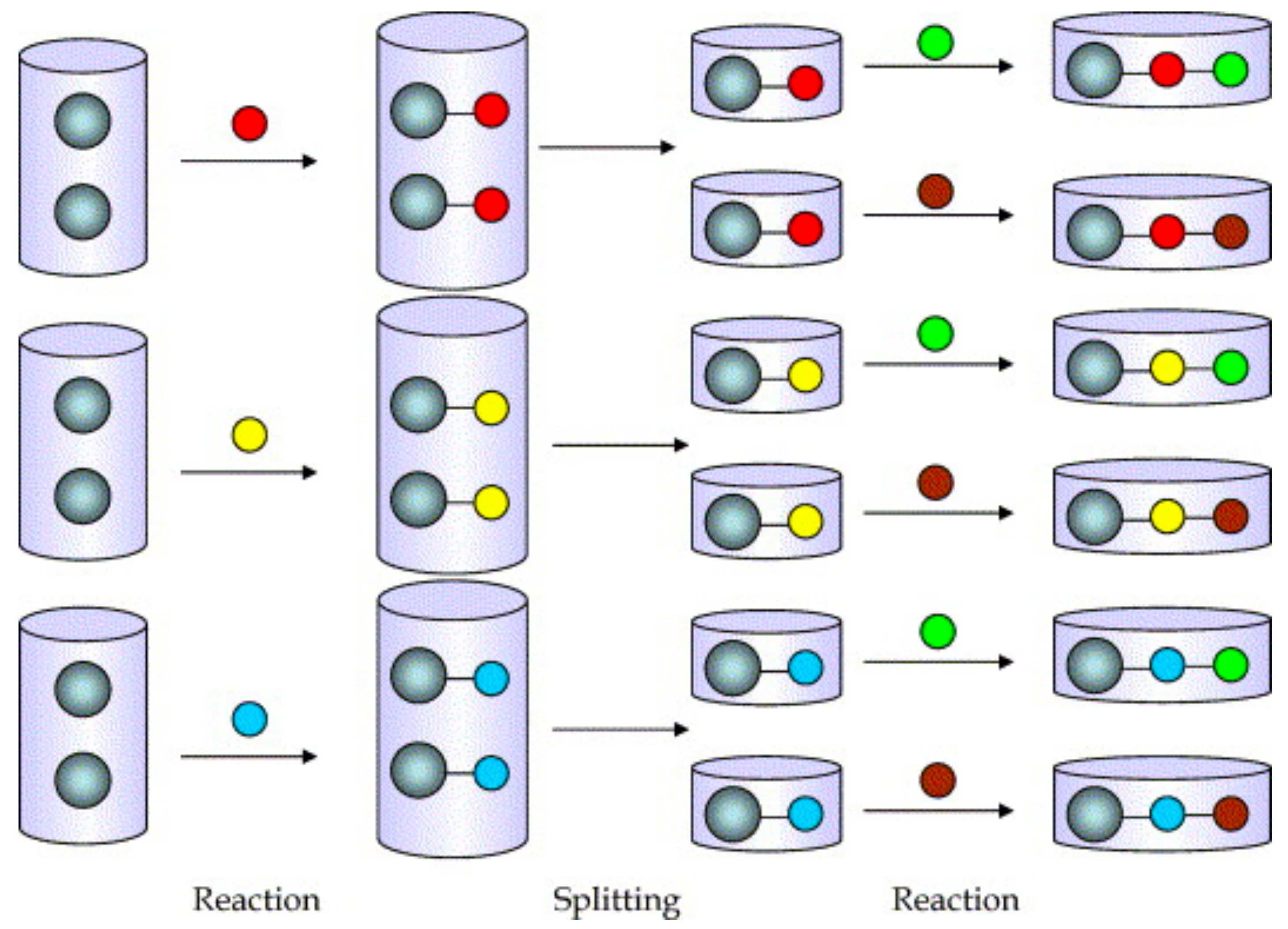
Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

Combinatorial Chemistry



We are developing solutions to generate massive quantities of training data called

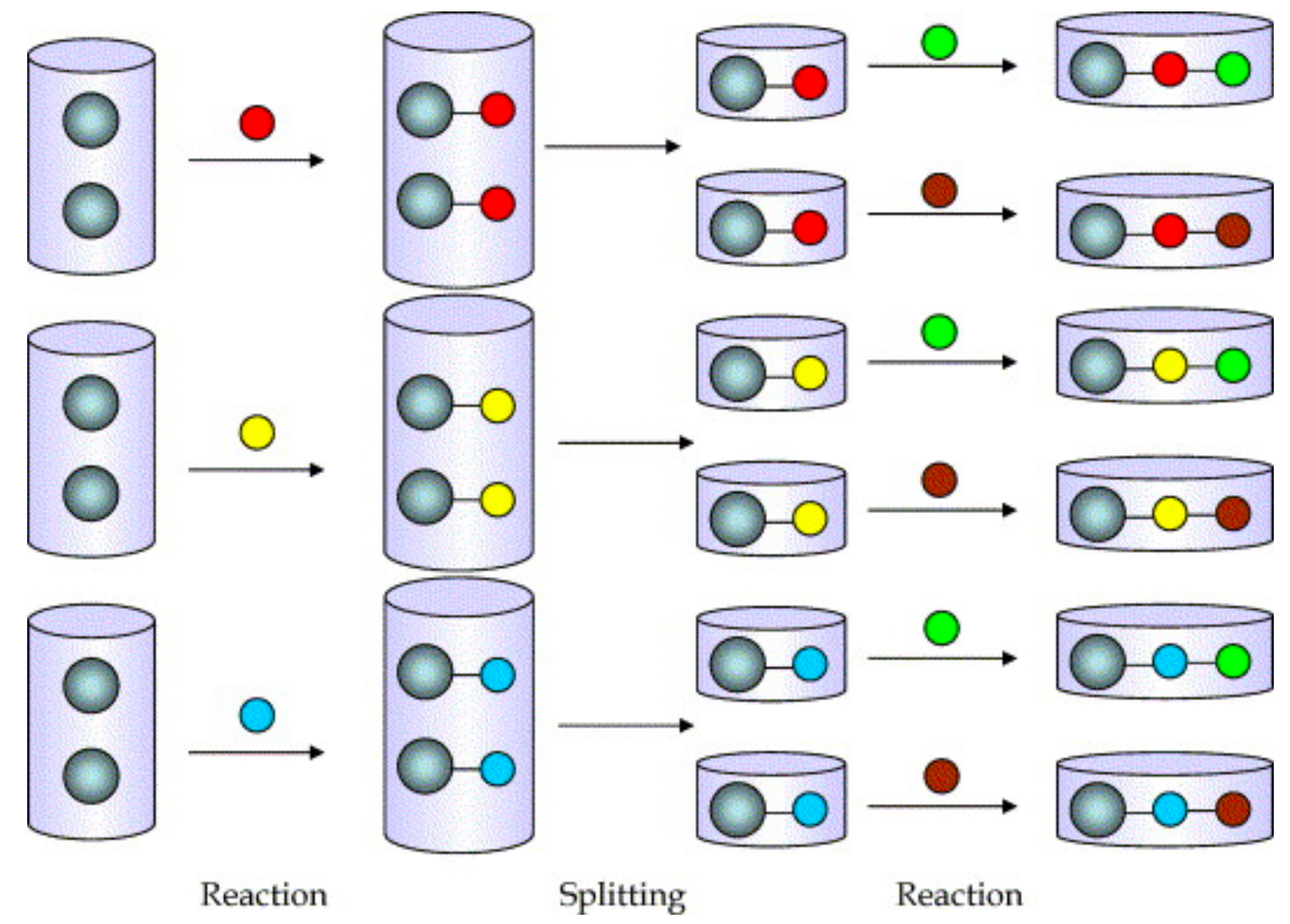
Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

Combinatorial Chemistry



Generate 10k **known** reaction products in a day for $5k

We are developing solutions to generate massive quantities of training data called

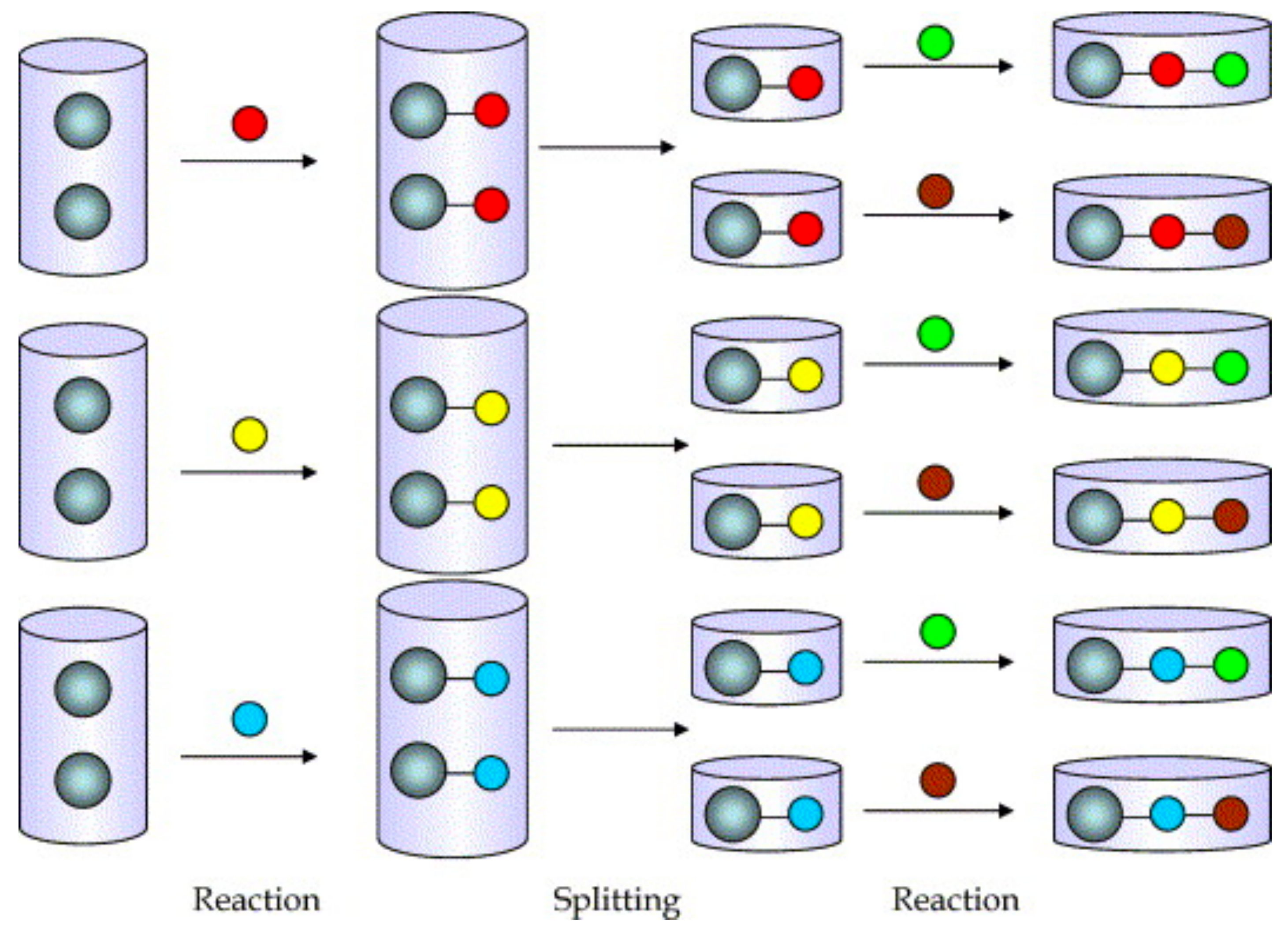
Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

Combinatorial Chemistry



Generate 10k **known** reaction products in a day for $5k

We are developing solutions to generate massive quantities of training data called

"Shotgun spectrometry"

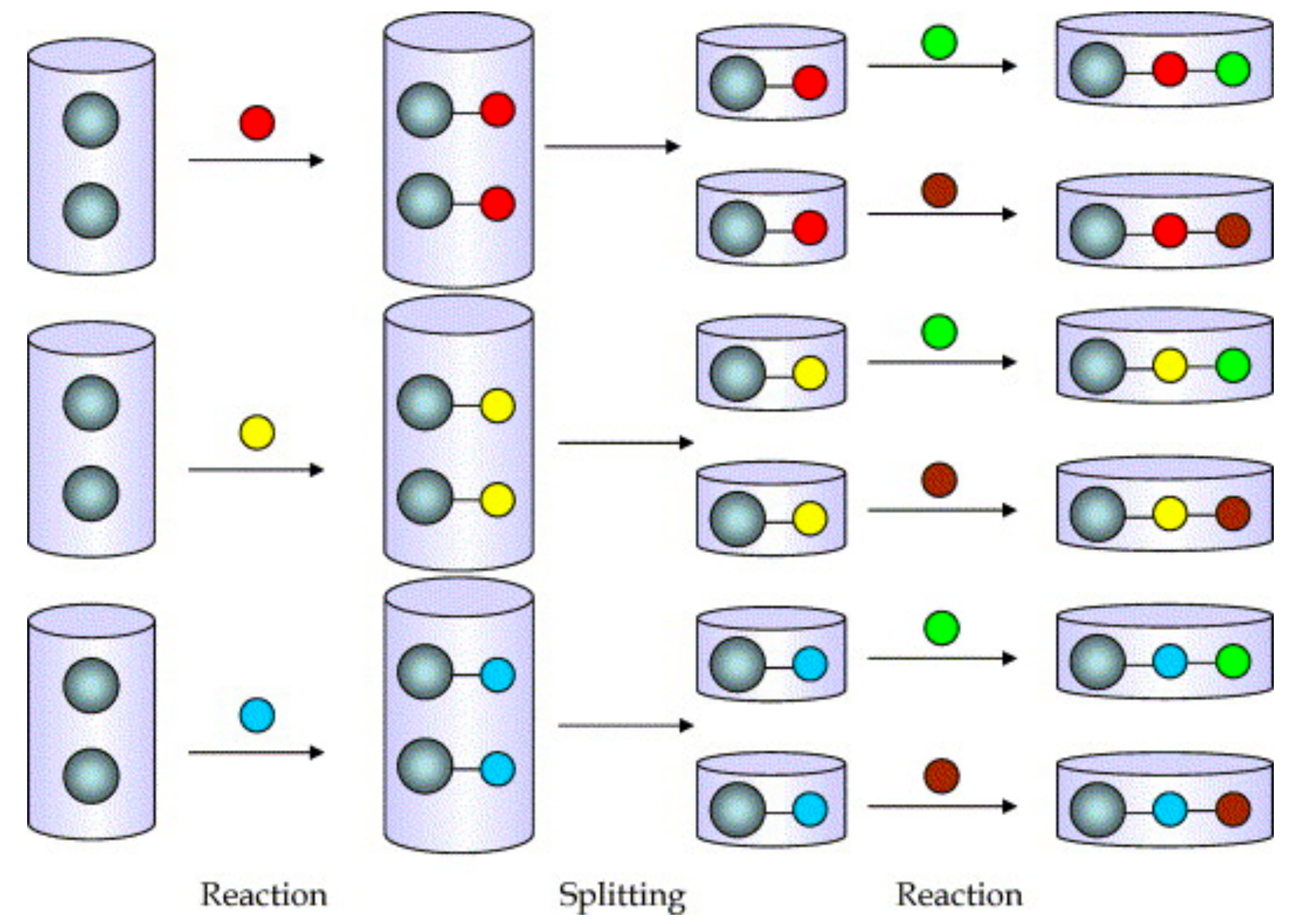
Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

Combinatorial Chemistry



Generate 10k **known** reaction products in a day for $5k

We are developing solutions to generate massive quantities of training data called

“Shotgun spectrometry”

1. Generate a mixture of 10k molecules with maximal molecular weight diversity
2. Perform LC/MS/MS on the combined mixture at each level
3. Train our new models on the resulting lightly-separated mixtures — some overlap will exist **and that’s ok**

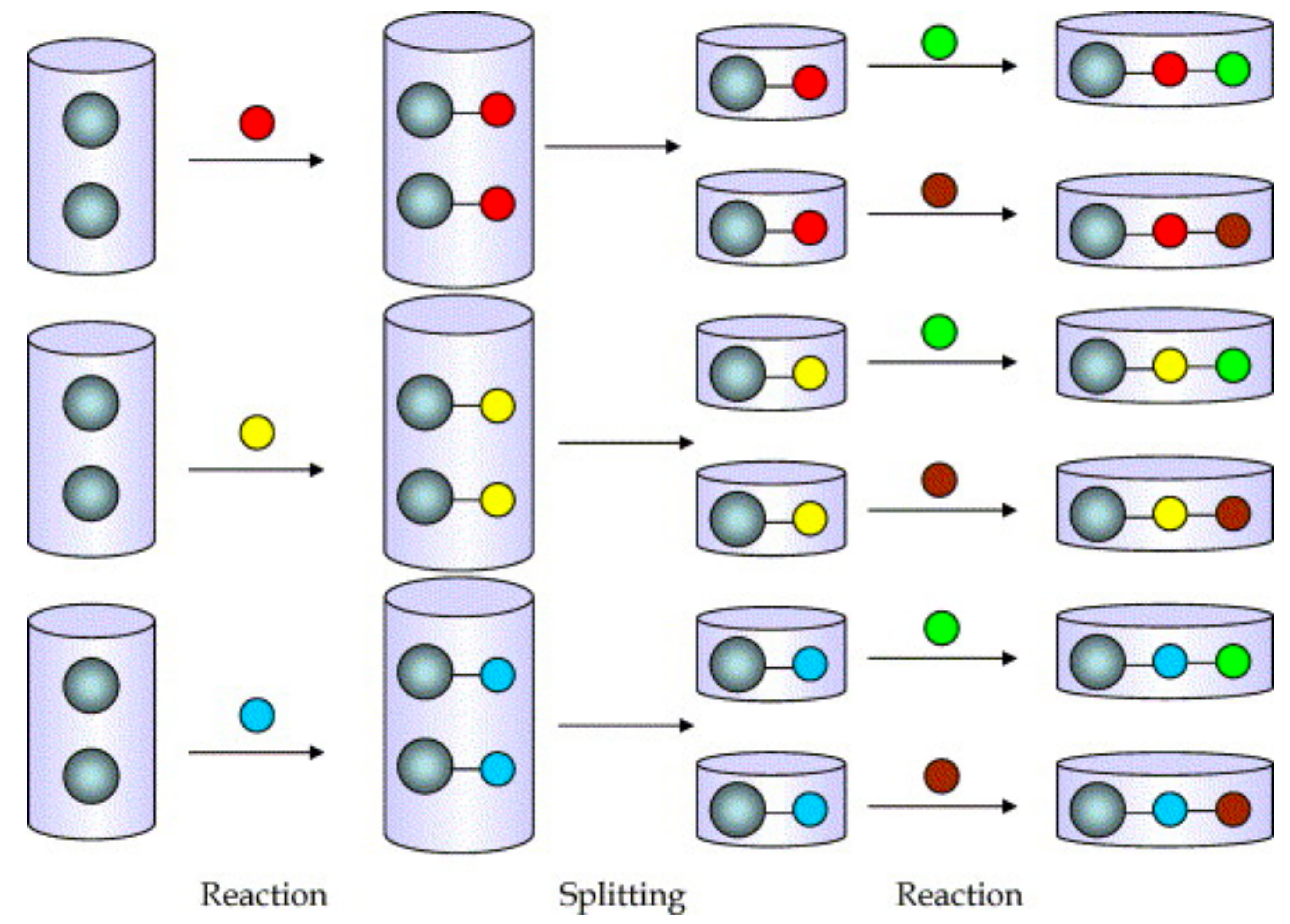
Completely solve MS forward model for $100k

Developing new techniques for acquisition

Modern AI requires tons of data

A modern MS machine can select a single m/z **before** fragmentation — useful for complex mixtures

Combinatorial Chemistry



Generate 10k **known** reaction products in a day for $5k

We are developing solutions to generate massive quantities of training data called

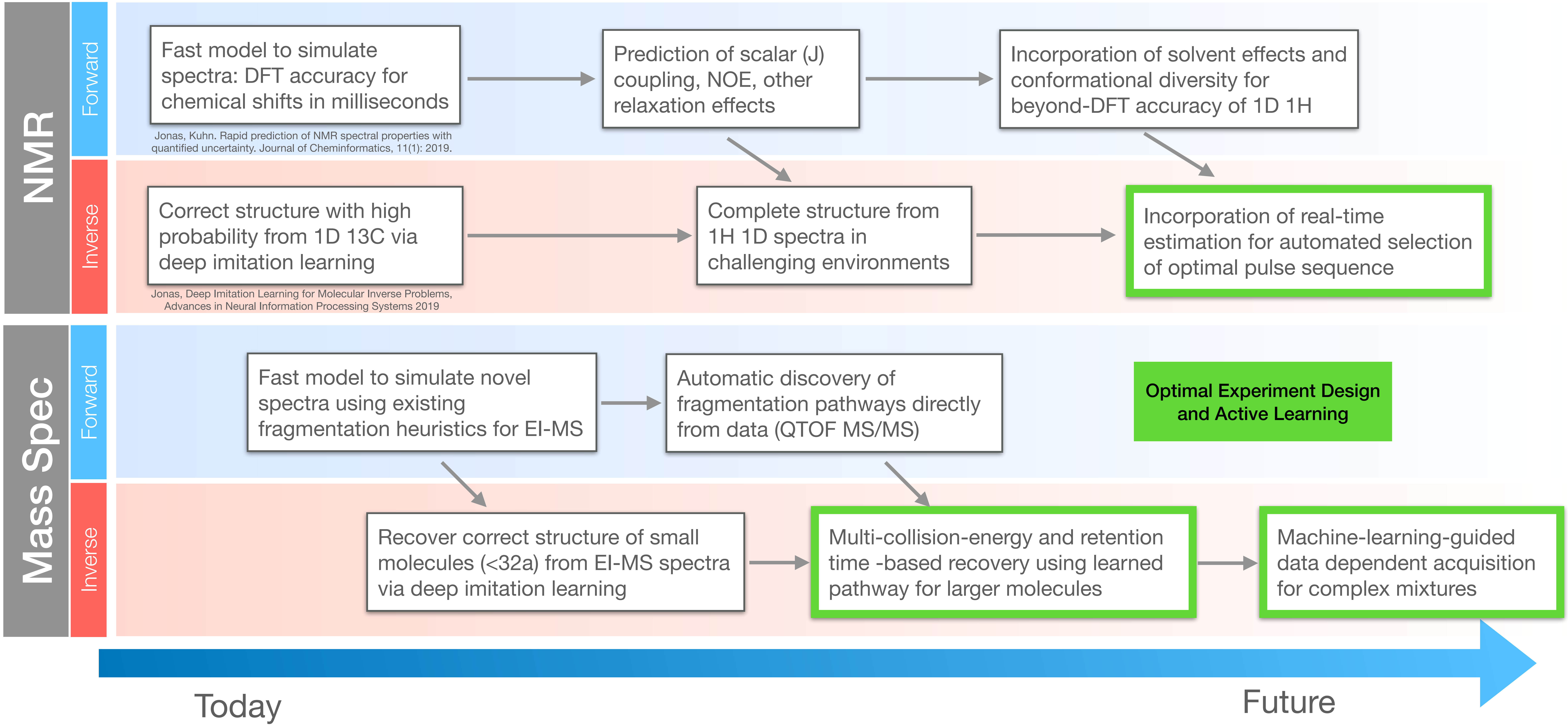
"Shotgun spectrometry"

1. Generate a mixture of 10k molecules with maximal molecular weight diversity
2. Perform LC/MS/MS on the combined mixture at each level
3. Train our new models on the resulting lightly-separated mixtures — some overlap will exist **and that's ok**

Can potentially scale CombChem to 100k per batch

Completely solve MS forward model for $100k

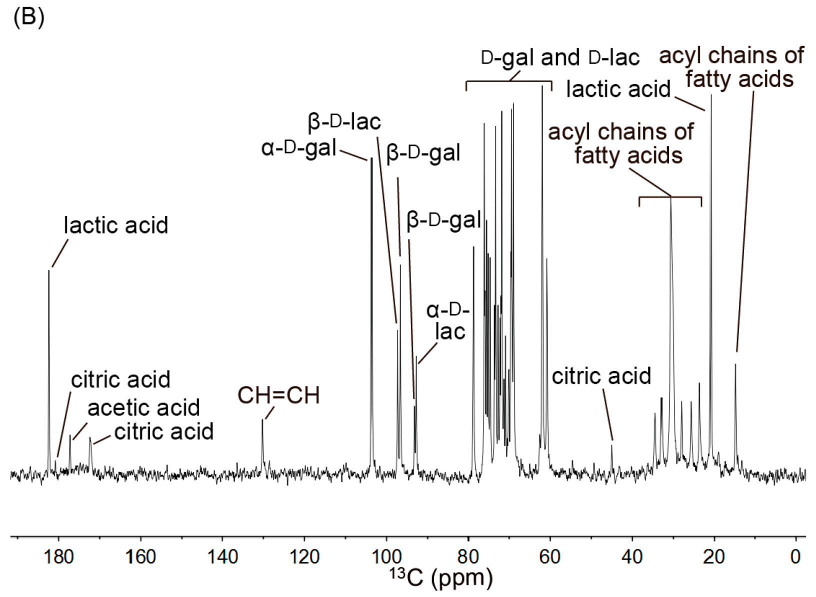
Research Plan



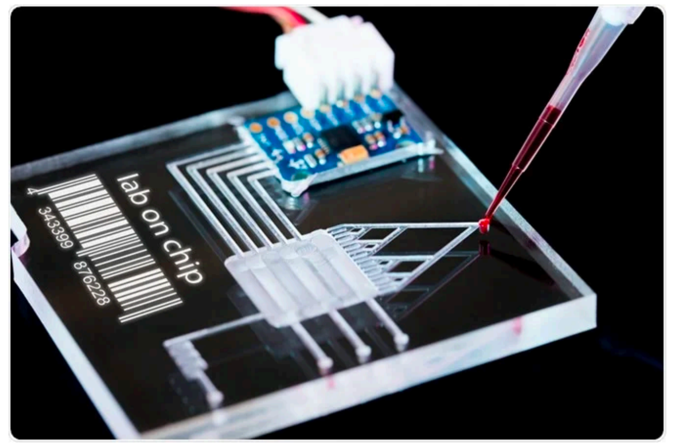
Self-driving spectrometers

Designing Algorithms, Software, and Systems to Measure Every Molecule

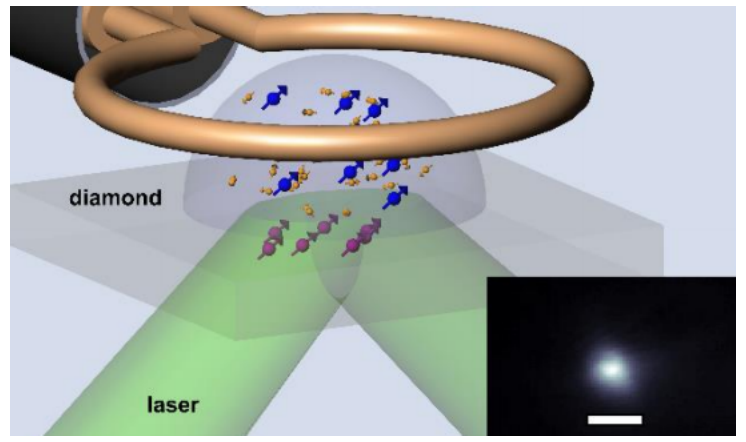
Making possible:



Scalable measurement of complex mixtures

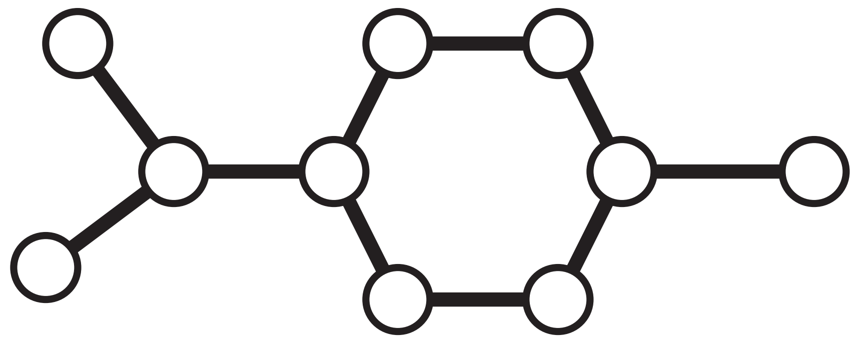


Allowing end-to-end laboratory automation

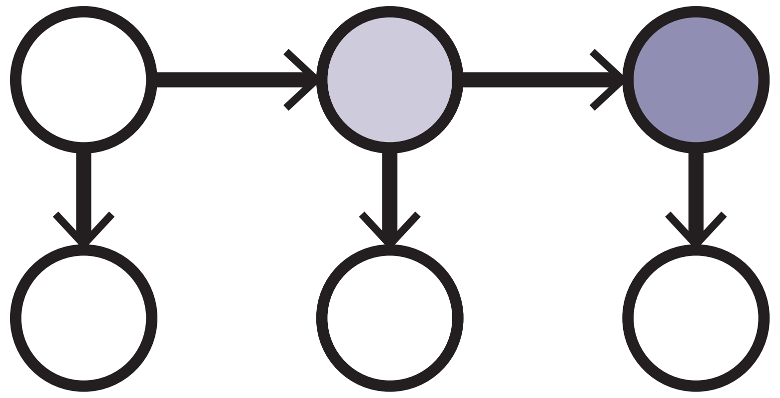


Ultimately allowing novel spectroscopic techniques

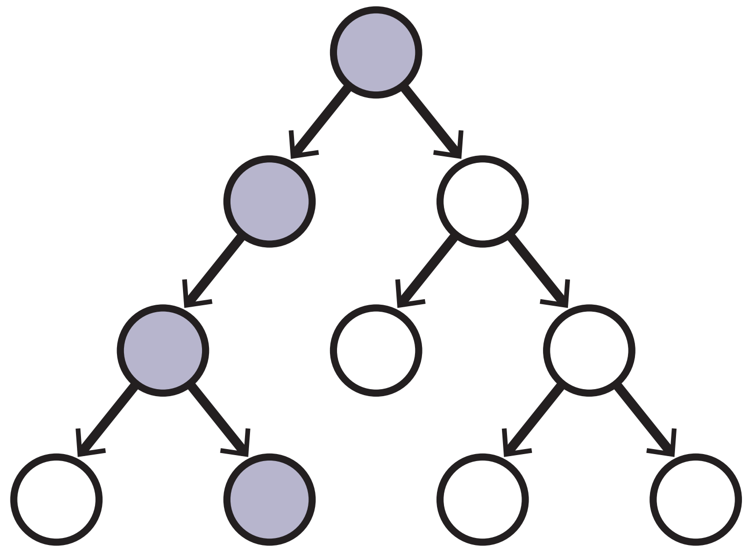
By new AI techniques:



Better fast forward models via physics-informed Graph NNs



Structured prediction via Deep Imitation Learning



Optimal experimentation via real-time active learning