Using Helmholtz Free Energy to Reduce Machine Learning Complexity

CASIS, 2018

Gerald Friedland (ENG, UC Berkeley), J. Wang (UC Berkeley), R. Jia (UC Berkeley), Bo Li (UC Berkeley), Dawns Song (UC Berkeley)

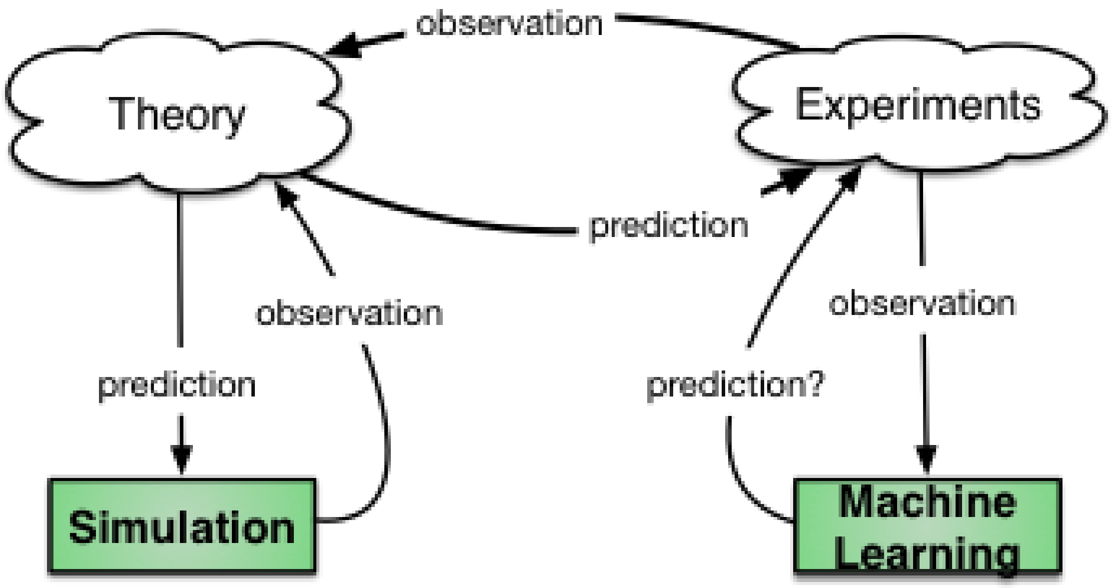
May 23, 2018

LLNL-PRES-751720

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

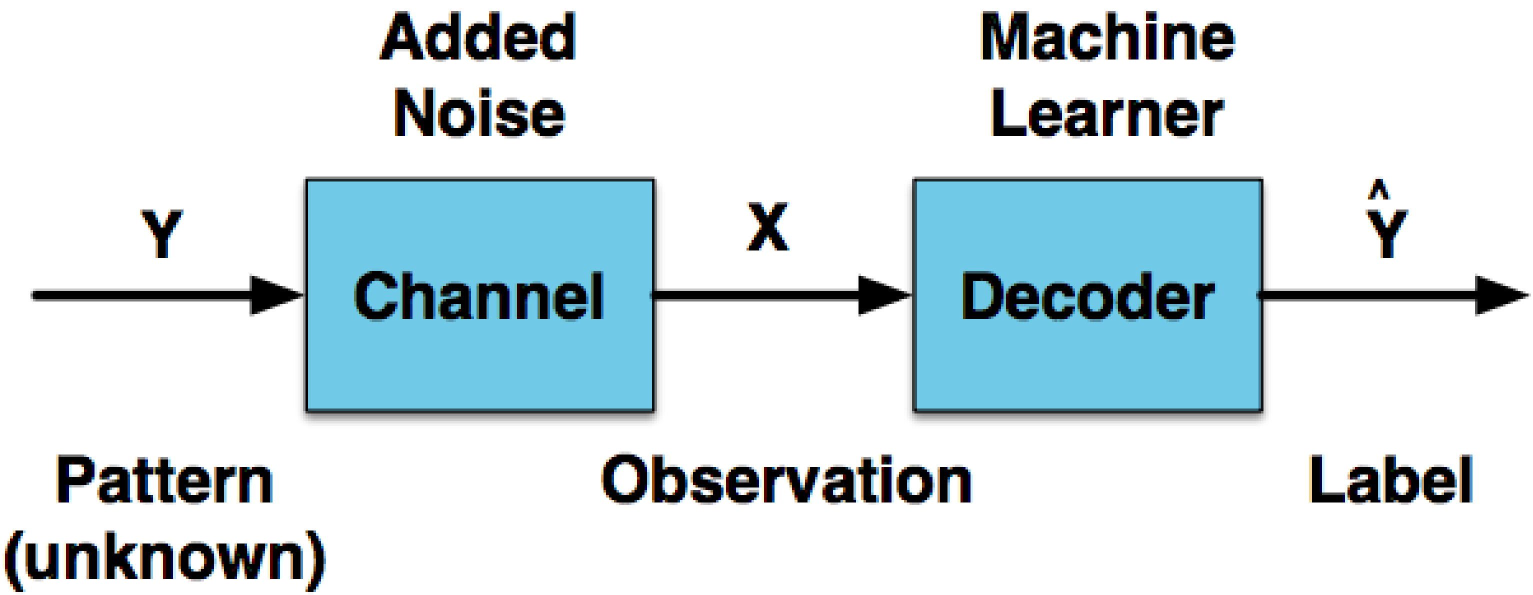
Background

- CASIS 2017: Information is Energy
- See also: R. Feynman's "Lectures on Computation" (1988)



Information Theory and Signal Processing make Machine Learning (ML) more efficient!

A Thermodynamic/Information Model for ML



- **Machine Learning resets bits introduced by noise.**
- **Machine Learning denoises an unknown pattern.**

Helmholtz free Energy

$$A \equiv U - TS,$$

- A= Free Energy
- U = Internal Energy
- T = Temperature
- S = Uncertainty

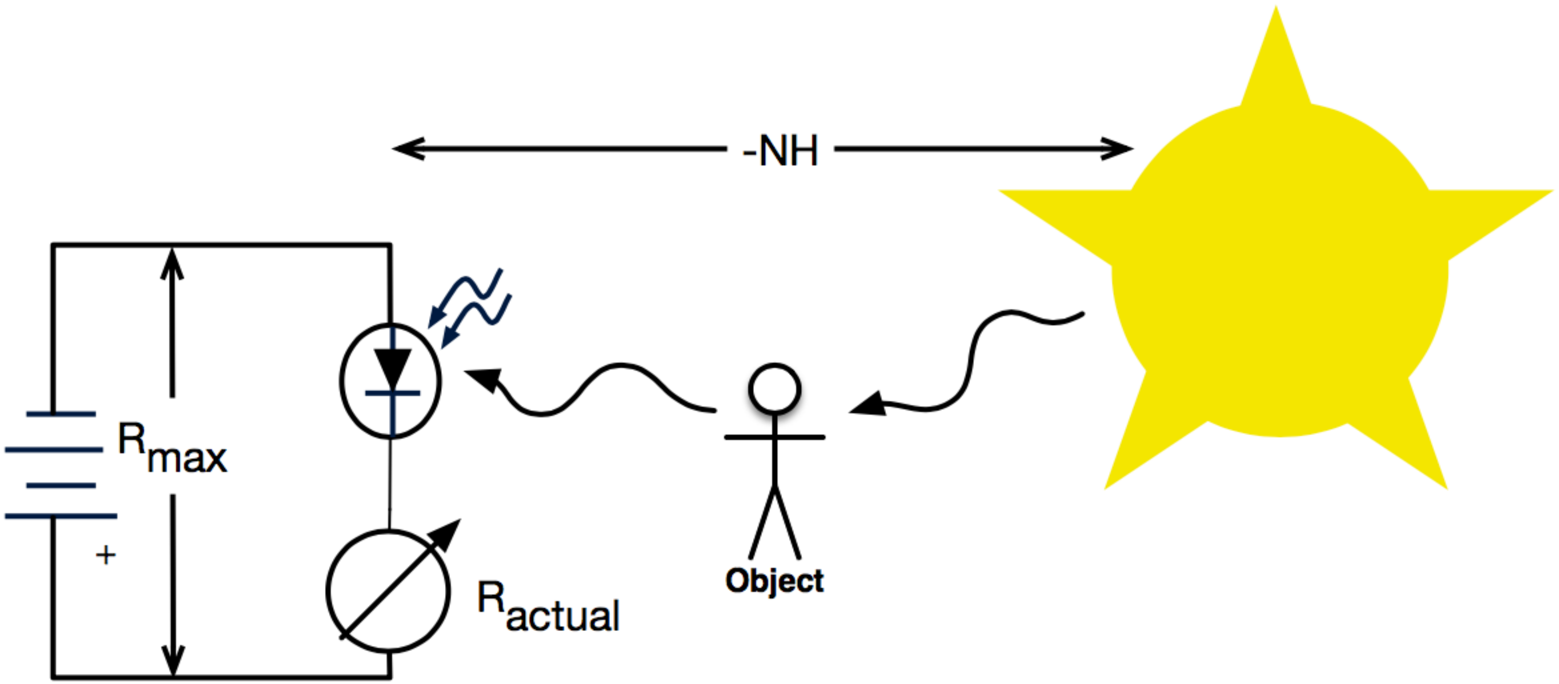
Shannon Entropy and Thermodynamic Entropy

$$H = U - S$$

Information is Reduction of Uncertainty

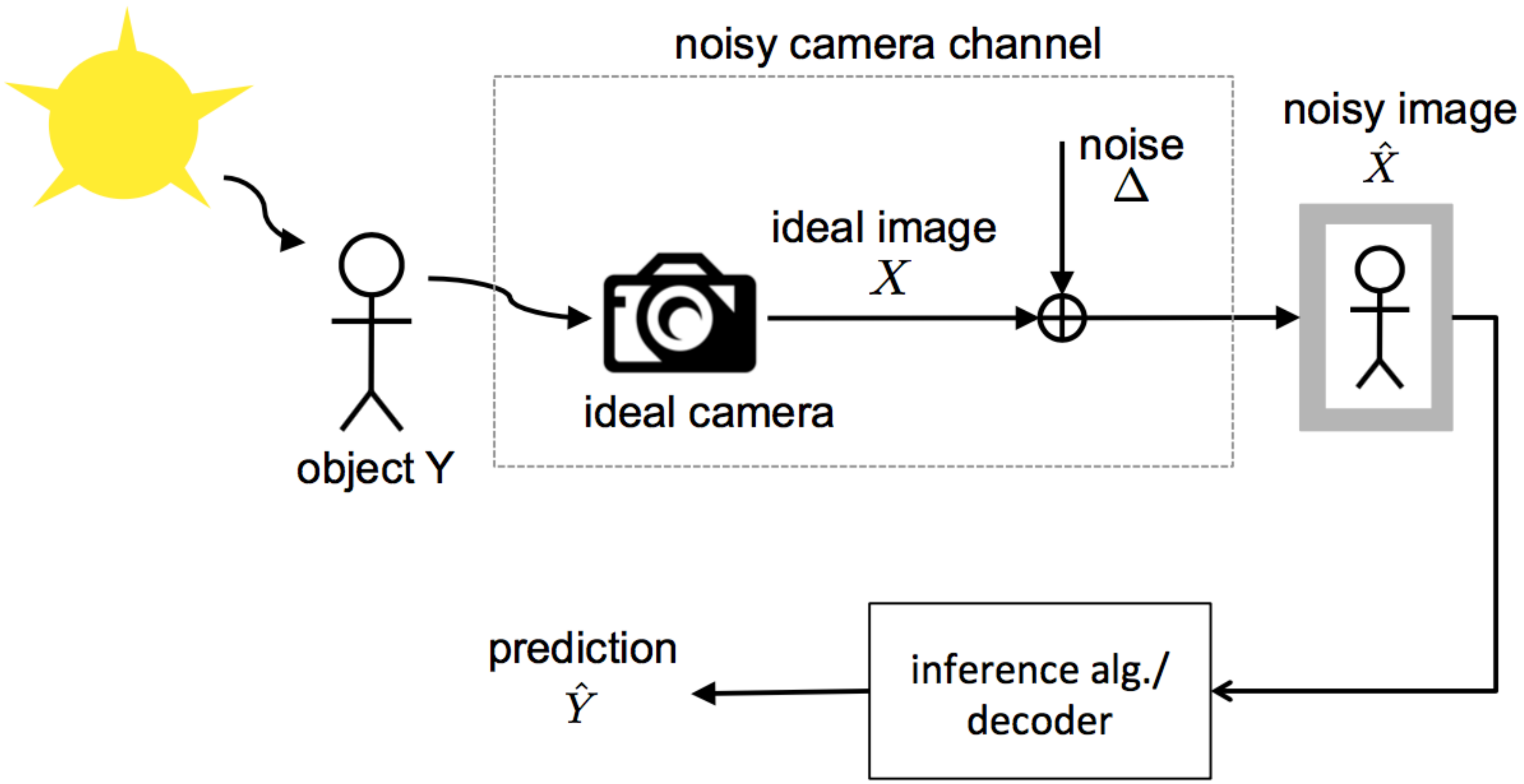
See also: Computation, Data and Science
https://www.youtube.com/playlist?list=PL17CtGMLr0Xz3vNK31TG7mJIzmF78vsFO

Reinterpretation



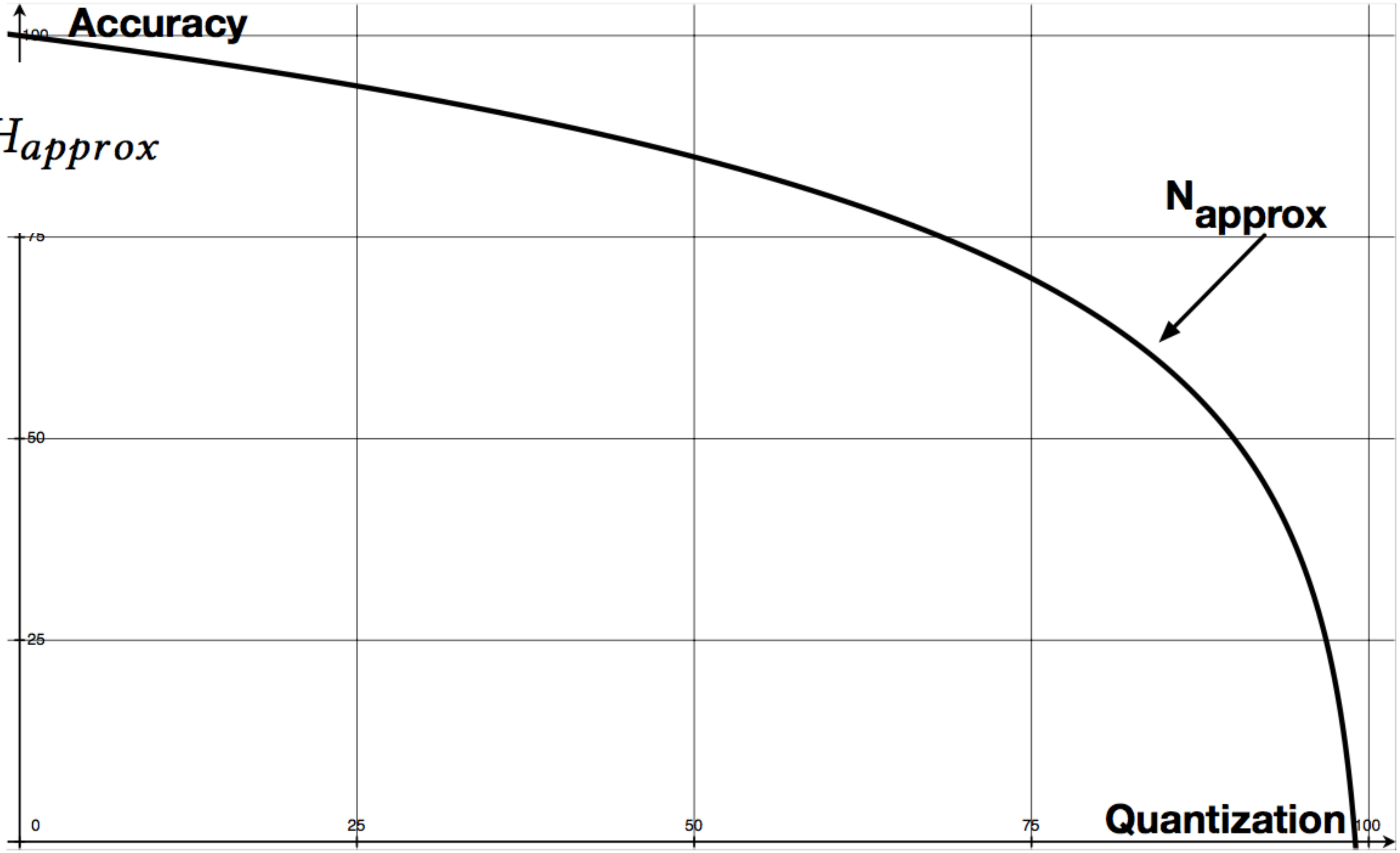
$$R_{actual} = R_{max} - NH$$

Reinterpretation with Information Theory



How does lossy compression work?

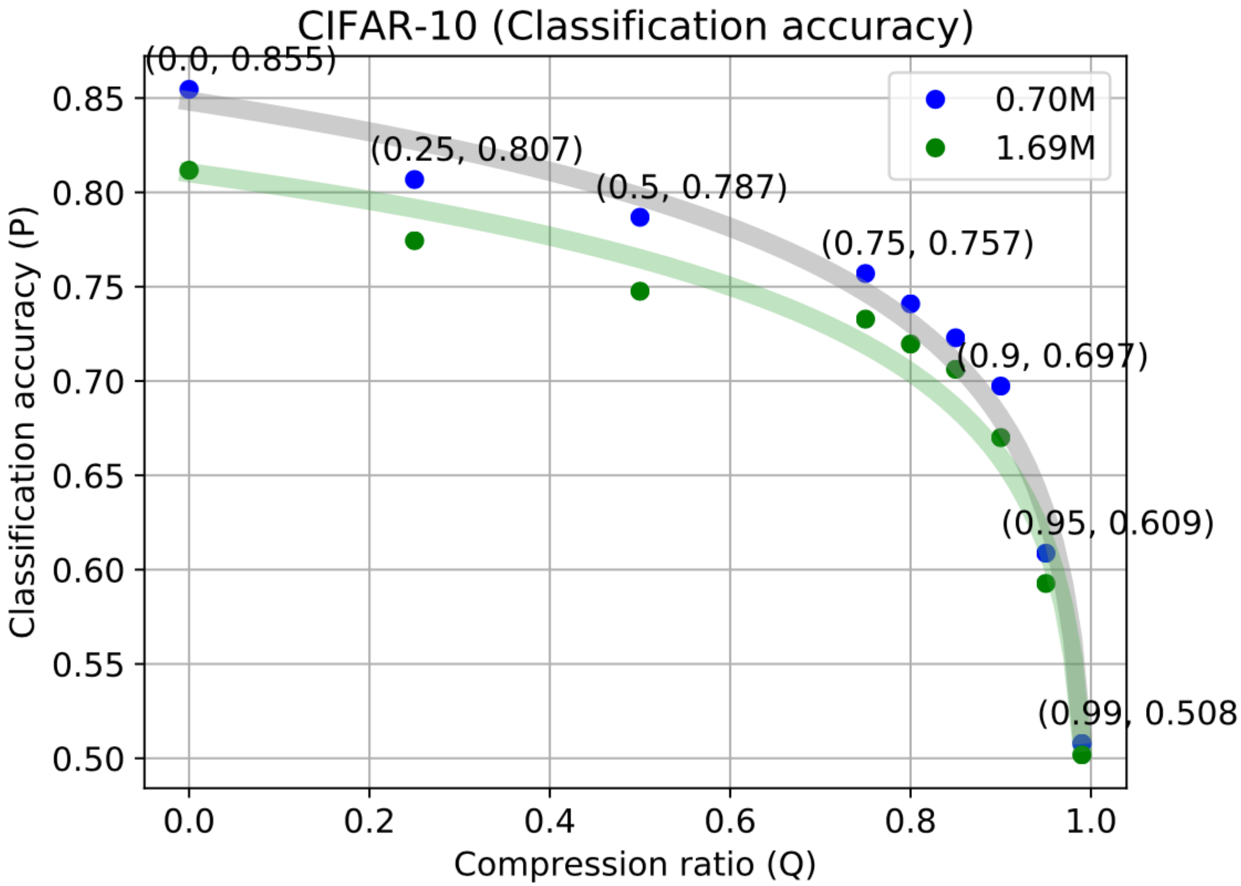
$$\frac{NH}{N_{approx}} = H_{approx}$$



Experiments: Images (overall)

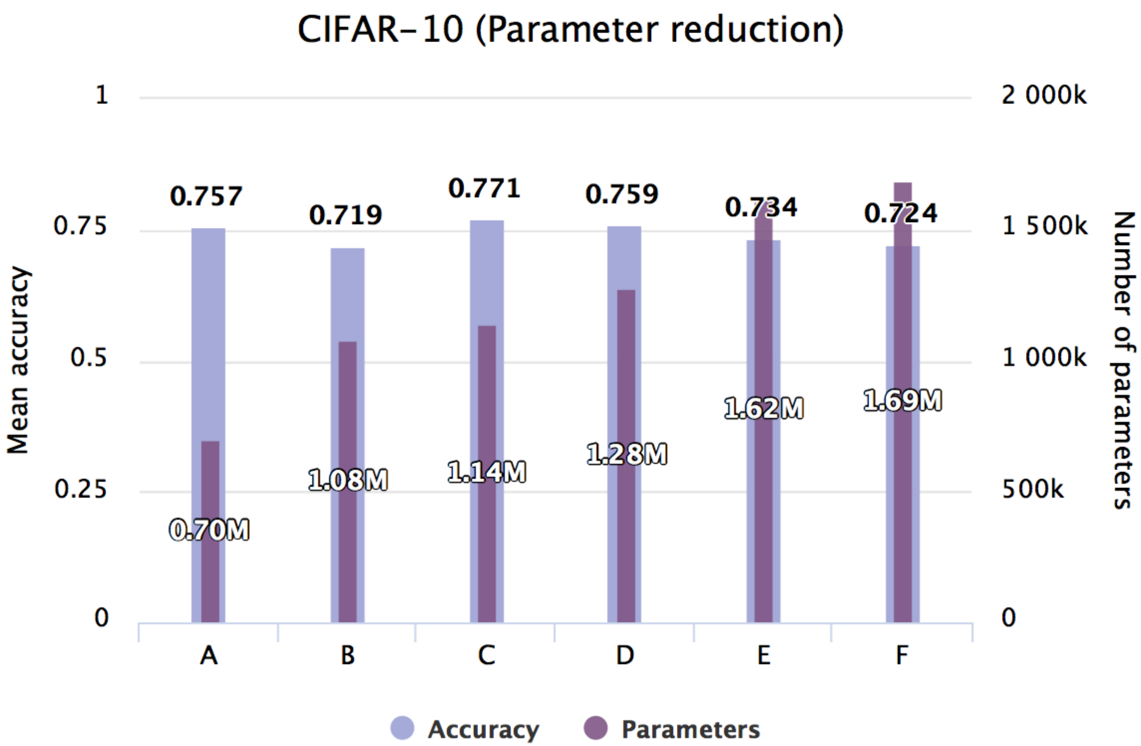
A	C	F
Conv([32, 64], 3, 3) + ReLU	Conv([32, 64], 3, 3) + ReLU	Conv([32, 64], 3, 3) + ReLU
Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)
Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU
Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)
Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU	Flatten
Conv(10, 3, 3)	Conv(128, 3, 3) + Dropout(0.5)	FC(128) + Dropout(0.5)
Global_avg_pooling	Conv([128, 128], 3, 3) + ReLU	FC(256) + Dropout(0.5)
Softmax	Conv(10, 3, 3)	FC(256) + Dropout(0.5)
	Global_avg_pooling	FC(10)
	Softmax	Softmax
701,386 (0.70M)	1,144,138 (1.14M)	1,686,090 (1.69M)

Experiments: Images (overall)



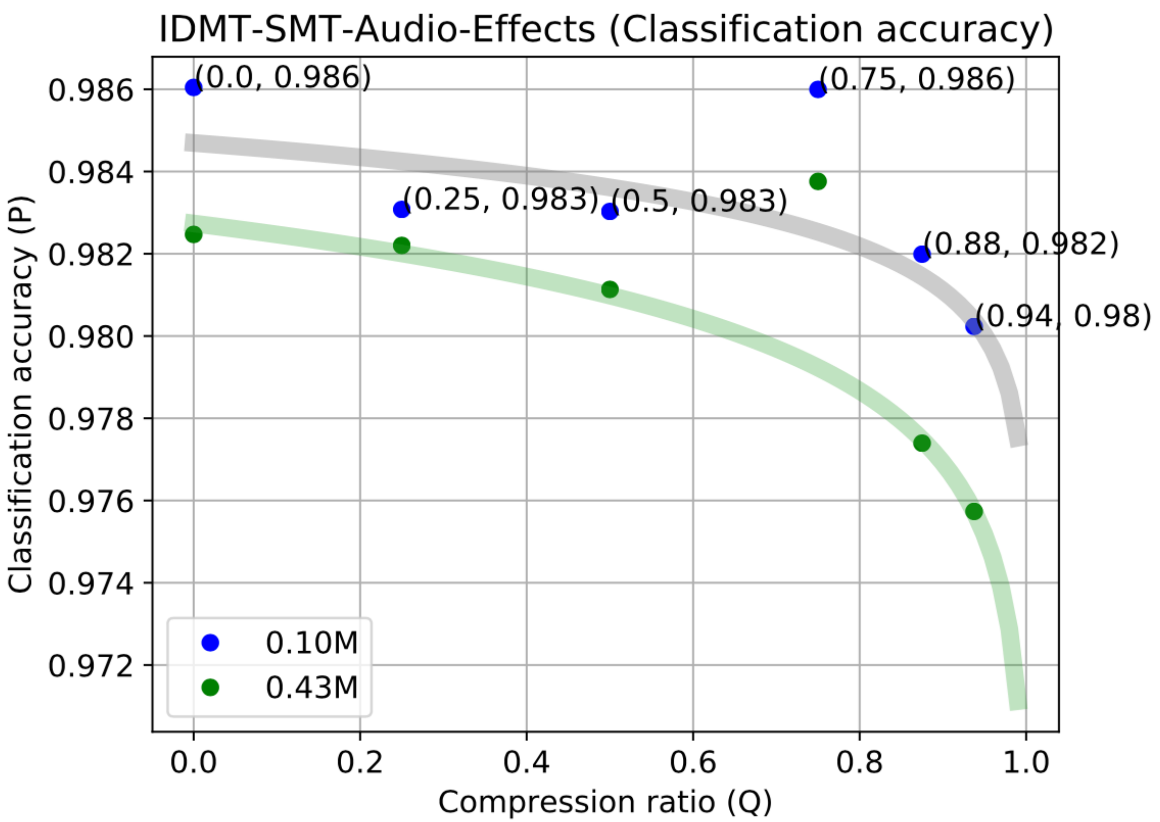
No Free Lunch!

Experiments: Images concrete



Less Parameters = Higher Accuracy!

Experiments: Audio



Experiments generalize to audio

Analysis: Images

JPEG quantization matrizes:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	36	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Best quality/accuracy trade-off (N_{approx}) around q=20.
This is at 1 bit/pixel!

Conclusion

- Sensor readings don't just appear. There is physics behind it.
- Understanding and using thermodynamics cuts down on parameters for machine learning and on search space (exponentially!)
- Example: We can discard about 23bits/pixel!
- Not presented here:
  - Adversarial examples are caused by redundancies
  - Thermodynamics of Computing helps understand Machine Learning