Machine Learning Geometry and String Theory

Jim Halverson
Northeastern University

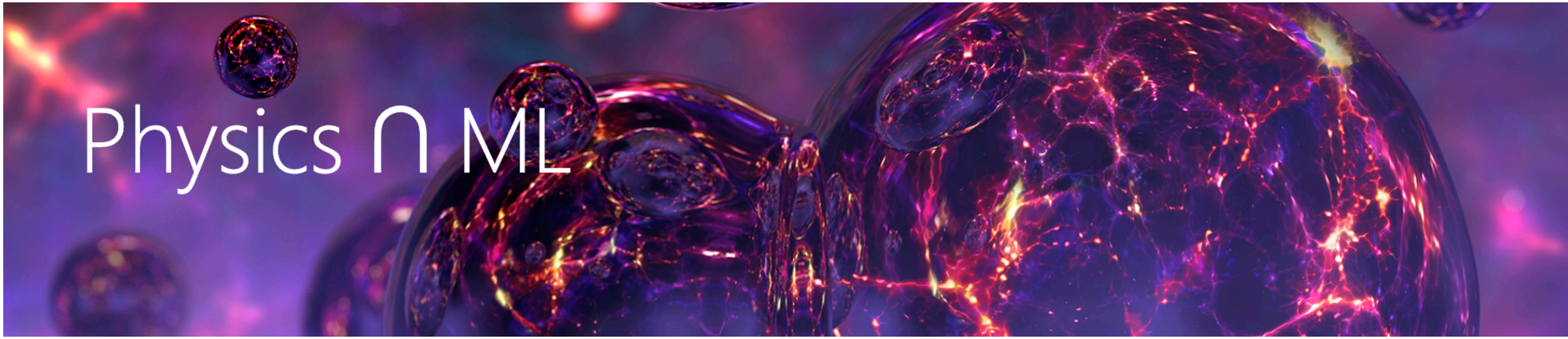
@ KITP, Crossroad of Physics and ML

Physics works with data science techniques:
1707.00655 with Jon Carifio, Dima Krioukov, and Brent Nelson
1711.06685 with Jon Carifio, Will Cunningham, Dima Krioukov, Cody Long, and Brent Nelson
1902.xxxxx with Nelson, Ruehle
1903.xxxxx with Long, Nelson, Salinas
19xx.xxxxx with Long, Ruehle, Tian

A large ensemble and universality:
1706.02299 with Cody Long and Ben Sung
1710.09374 with Cody Long and Ben Sung

Computational Complexity and Undecidability:
1809.08279 with Ruehle
1009.5386 with Cvetic, Garcia-Etxebarria

Microsoft | Research



Overview

Region: North America

Date: April 25, 2019 - April 26, 2019

Location: Redmond, Washington, USA

Venue: Microsoft Research Building 99/1919
Redmond, Washington, USA

About Agenda

The goal of *Physics ∩ ML* (read 'Physics Meets ML') is to bring together researchers from machine learning and physics to learn from each other and push research forward together. In this inaugural edition, we will especially highlight some amazing progress made in string theory with machine learning and in the understanding of deep learning from a physical angle. Nevertheless, we invite a cast with wide ranging expertise in order to spark new ideas. Plenary sessions from experts in each field and shorter specialized talks will introduce existing research. We will hold moderated discussions and breakout groups in which participants can identify problems and hopefully begin new collaborations in both directions. For example, physical insights can motivate advanced algorithms in machine learning, and analysis of geometric and topological datasets with machine learning can yield critical new insights in fundamental physics.

Please contact an organizing member if you wish to participate in this workshop.

Organizers

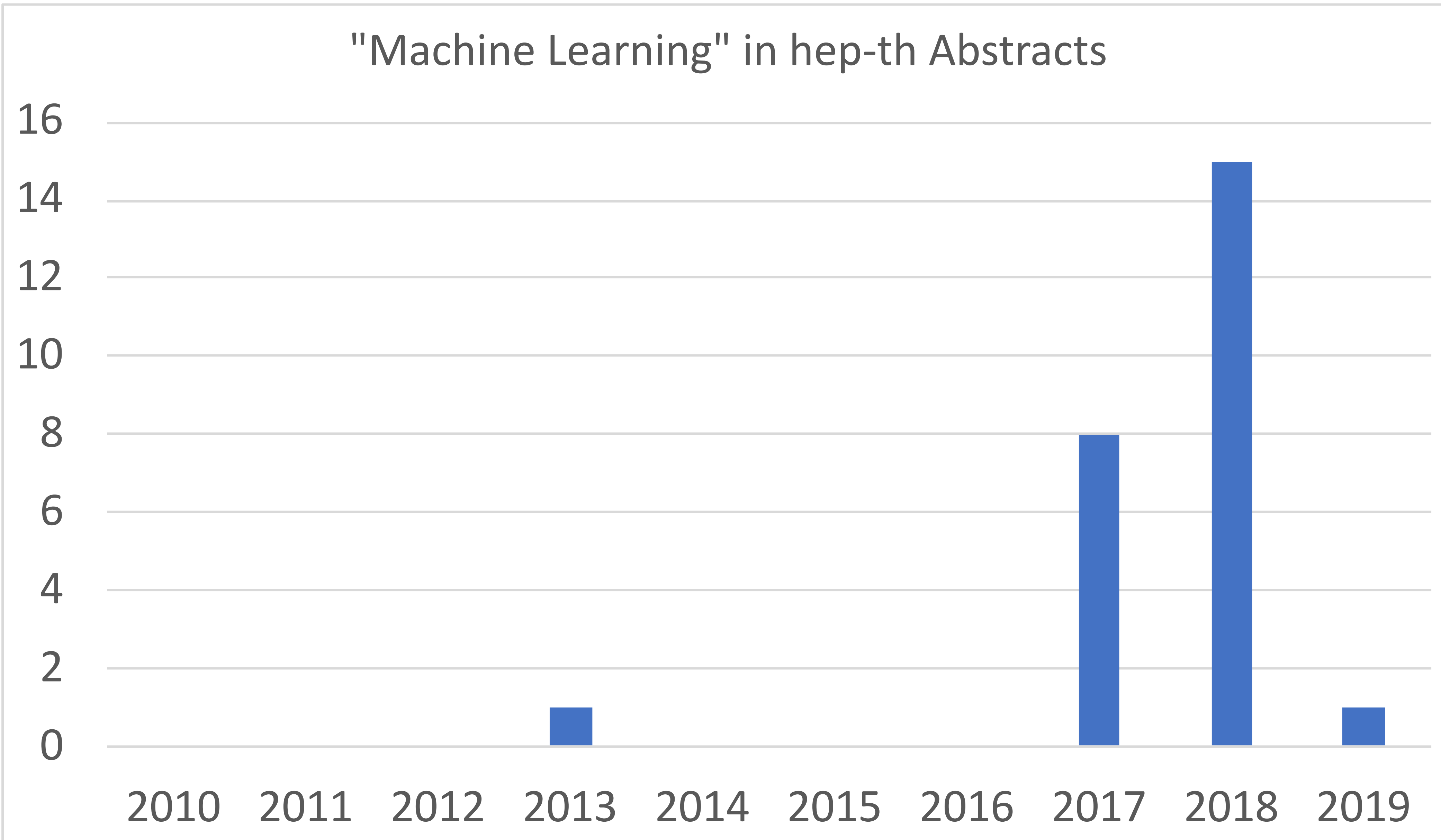
Greg Yang, Microsoft Research
Jim Halverson, Northeastern University
Sven Krippendorf, LMU Munich
Fabian Ruehle, CERN, Oxford University
Rak-Kyeong Seong, Tsinghua University
Gary Shiu, University of Wisconsin

Microsoft Advisers

Chris Bishop, Microsoft Research
Jennifer Chayes, Microsoft Research
Michael Freedman, Microsoft Research
Paul Smolensky, Microsoft Research

Application:

https://www.microsoftevents.com/profile/form/index.cfm?PKformID=0x5969440abcd



25 papers, more like O(30-40) with more inclusive terms.
O(10-15) works in progress, presented @ workshops.

Strings + ML @ 10,000 ft

Meetings:

- Northeastern, Nov 2017
- LMU Munich, March 2018
- TSIMF Sanya, June 2018
- ICTP Trieste, Dec 2018
- Microsoft Res, Apri 2019
- others in the works . . .

Scientific Directions:

- string vacua / landscape
- associated applications in particle physics + cosmology.
- algebraic geometry + math
- holography + QCD
- SUSY gauge theory

Techniques use so far:

- supervised learning (simple and various NNs)
- GANs
- Reinforcement learning
- Network Science
- Autoencoders
- Genetic algorithms
- others I probably forgot . . .

Selected Literature:

Evolving neural networks with genetic algorithms to study the String Landscape
Fabian Ruehle

Deep Learning and Holographic QCD
Koji Hashimoto, Sotaro Sugishita, Akinori Tanaka, Akio Tomiya

Machine Learning of Calabi-Yau Volumes
Daniel Krefl, Rak-Kyeong Seong

(apologies to many friends with excellent works not listed here, I wanted to keep the list short and across different areas!)

The problem:
Many questions.
Many (theoretical) data sets.

Today: fix one, address numerous physical questions in it.

A fun maximally conservative estimate:

$> 10^{742}$ TB to store.

$> 10^{684}$ T_{univ} to process.

Our big physics question:

is particle physics and cosmology described by string theory?

Our small physics question:

how do string ensembles change our particle-cosmology perspective relative to the bottom-up?

Upshot: landscape —> complex system

Outline

- String landscape 101. (sketch physics, size, complexity).
- **Today's data:** ensemble of topologically distinct transition-connected geometries with physics universality.
- ML application 1: **Q: Where is something like the SM?** simple techniques —> conjecture —> gauge sector theorem.
- ML application 2: **Q: Where does this data set "fit"?** understanding control by probing weak coupling with RL.
- ML application 3: **Q: Why are there SM particles?** early universe reheating and simplifications from SL.

Strings at Low Energies

- Theory of quantum gravity

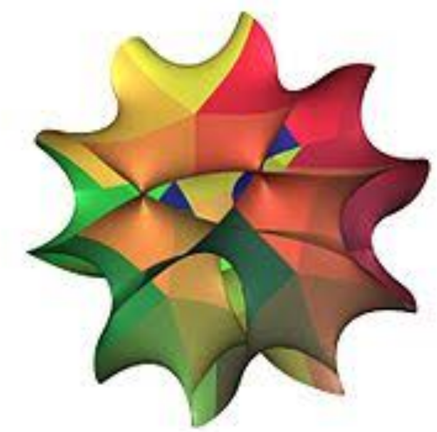
- Extra dimensions:

  - have structure (e.g. Calabi-Yau)

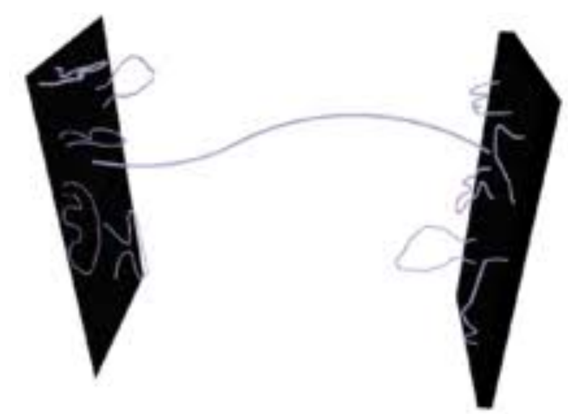
  - can have things in them
    (e.g. branes, fluxes)

- There are solutions, each with EFT

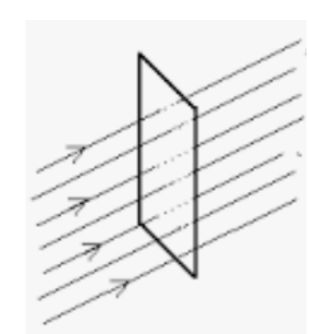
- But many solutions!
  The string landscape



+



+



= $\mathscr{L}$

Landscape Size

- Early result: $O(10^{500})$ weakly coupled IIB flux vacua
  [Ashok, Douglas] [Denef, Douglas]

- Update: $O(10^{272,000})$ F-theory flux vacua. (one geometry)
  [Taylor, Wang]

- Geometries exact lower bound: 10^{755} (today's data)
  [JH, Long, Sung]

- Believed to be finite. Some examples in corners:

  - consistency + SUSY in brane systems:
    [Douglas, Taylor] [Cvetic, JH, Klevers, Song]
  - 6d F-theory elliptic fibrations: [Gross], [Grassi]
  - 4d F-theory elliptic fibrations: [DiCerbo, Svaldi]

Landscape Complexity

- **Seminal work:** small cosmological constants are NP-complete by reduction from subset sum. [Denef, Douglas]
- **Undecidability:** tremendous number of diophantine problems, both from decision problems and index theorems. Patchy landscape! [Cvetic, JH, Garcia-Etxebarria]
- **Cosmological dynamics:** complexity effects on vacuum selection. [Denef, Douglas, Greene, Zukowski]
- **Fast optimization for toy (ADK ~ knapsack) cosmological constants:** [Bao, Bousso, Jordan, Lackey]
- **Finding vacua:** computing scalar potential in strings and finding minima in associated EFT are both hard. [JH, Ruehle]

Today's Data

- *topologically distinct geoms (6-manifolds)*
- *all connected by topological transitions.*
- *have some universal physical features*

[JH, Long, Sung] x 2, 1706 and 1709

The Mathematics

- **4D F-theory:** 3-fld B, gauge structure det'd by B topology, called "non-Higgsable cluster."

Some selective progress: Anderson, JH, Heckman, Grassi, Morrison, Rudelius, Shaneson, Taylor, Wang, Vafa.

- **Starting point:** B a weak Fano toric threefold, encoded in a certain triangulation (fine regular star) of a 3d reflexive polytope.
- **Topological transitions:** systematically perform sequences of "toric blowups" — topological surgery along two-spheres, points.
- **Sequences are bounded:** if all singularities are "canonical", geom. is at finite distance from bulk of field space in the Weil-Petersson metric.

Alg. Geom: [Hayakawa] [Wang] **in F-theory: [Morrison]**

- **Classification:** there are 82 (41,873,645) sequences over curves (points) that satisfy a sufficient condition for canonical singularities.
- **Ensemble:** all ways of performing these sequences of blowups. from an initial, fixed, triangulated polytope.

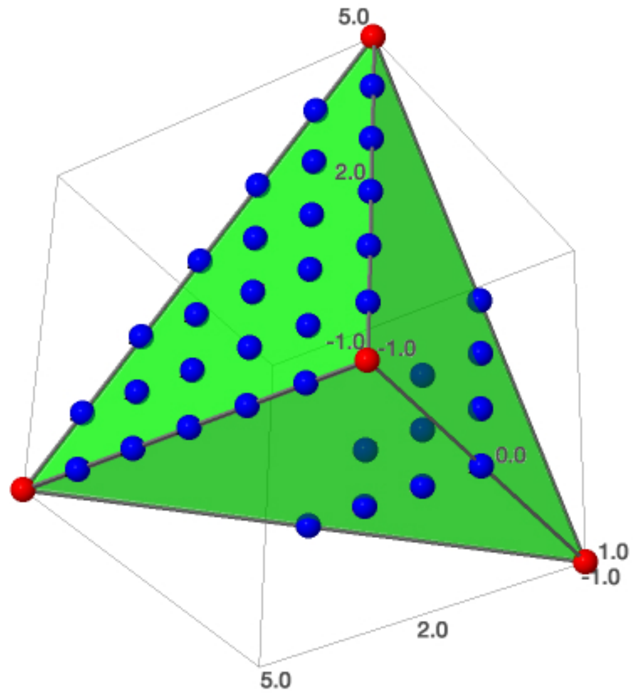
Fancy lingo = ?

Job security!

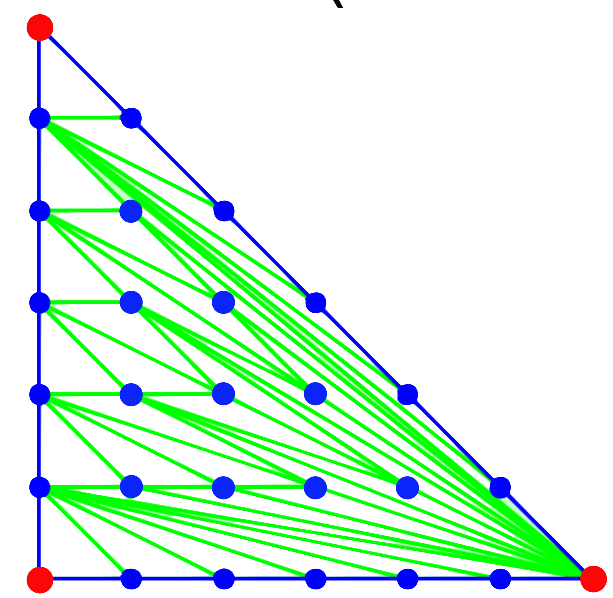
Reality: each geometry is specified by a constrained set of sets of vectors in $\mathbb{Z}^3$

The Simple Picture

- **Polytope:**

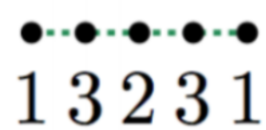


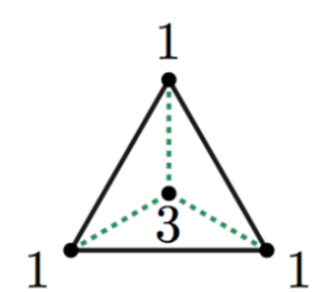
Triangulation: (codim 1 faces)

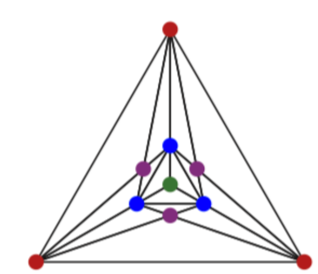


Fact: any FRS triangulation of this has 108 edges, 72 faces.

- **Rep seq. of blowups:** (topological transitions, project into board)







- **Ensemble Size:** (put the widgets on the triangulation)

$$82^{108} \times 41873645^{72} = 2.96 \times 10^{755}$$

The Integer

exact lower bound on topologically distinct F-theory geometries.

2962315213066588476735298783424968660803780217820767222414624251500652985088456211 55
695644076004524635389132997146097067111728163388031396056782083347614866566976655879
021984012793769157213779169816753325335978325856186523598065404800375507248464342388
511047235006655822119955724586948796896518847577026036417026840946433790956016724443
830558425085959438371548897609540032670520282917998350399793115958231423494399655849
101565013413100994870230661418306662241733727360211923208425845972788270516755913617
371792570199269731190754783864008989335957900232620795965279395190240666076073359738
714472109475716934497800719015826299673405211214139176513510727323709162299991669746
741171716096000000000000000000000000000000000000000000000000000000000000000000000000

of 4319 3d reflexive polytopes,
there’s one other polytope that yields this same number of geometries.
they dominate the ensembles from other polytopes
by over 60 orders of magnitude.

Analytic Physics Universality

related ensemble of [Taylor, Wang] has similar results

- **Universality from algorithm:** (nice when this possible)
  geometric ansatz with computable high prob. —> physics property

- for any geom., easy to compute geometric 7-brane structure at generic CS

- **Universality of Non-Higgsable Seven-branes:**

$$P(\text{NHC in } S_{\Delta_1^\circ}) \geq 1 - 1.01 \times 10^{-755}$$

$$P(\text{NHC in } S_{\Delta_2^\circ}) \geq 1 - .338 \times 10^{-755}$$

- **Universality of Large Gauge Sectors: prob > 99.9995 %** $rk(G) \geq 160$

$$G \geq E_8^{10} \times F_4^{18} \times U^9 \times F_4^{H_2} \times G_2^{H_3} \times A_1^{H_4} \qquad U \in \{G_2, F_4, E_6\}$$

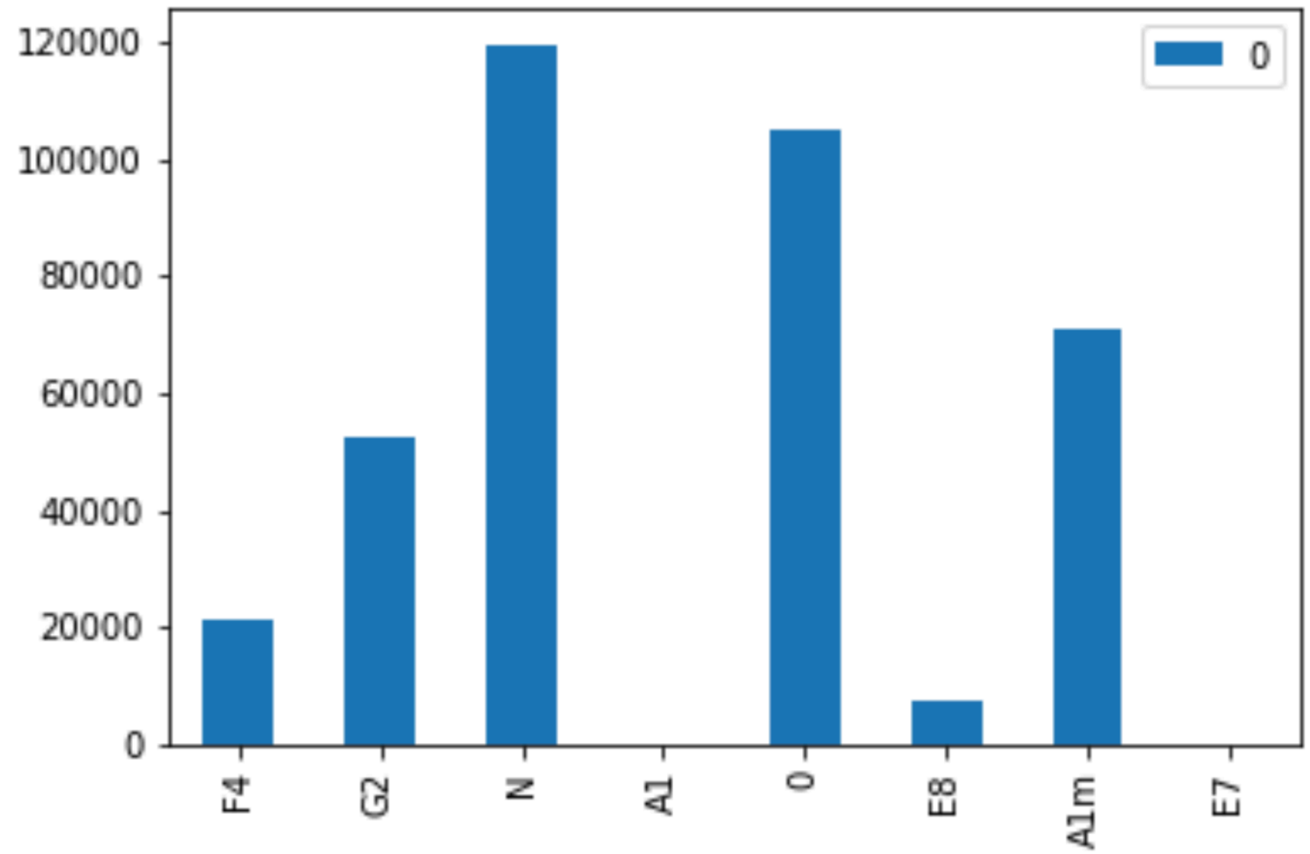
$$rk(G) \geq 160 + 4H_2 + 2H_3 + H_4$$

- **Cosmology Suggestion: Dark Glueballs?**

- **Universality of Strong Coupling:** $\frac{N_{\text{Sen}}}{N_{\text{Total}}} \leq 3.0 \times 10^{-391}$

Better Universality from Sampling

from 200 random samples



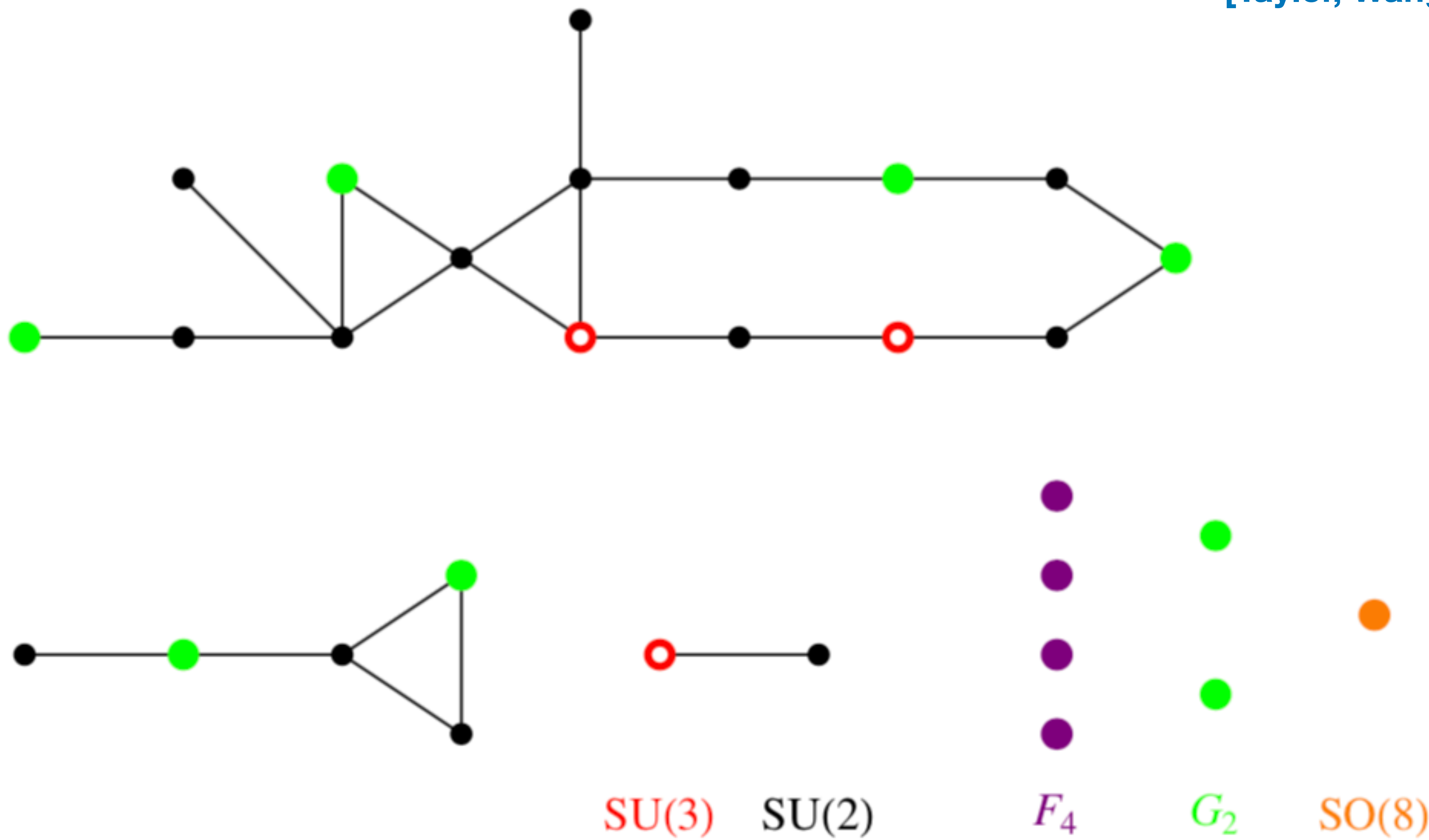
- avg # axion-like particles (ALPs): 1883 +- 29. **(Experiments!)**

- avg # ALPS that are gauge axions = avg # of gauge factors: 762 +- 11

- avg rk of gauge group: 1609 +- 17.

“Typical” Example

[Taylor, Wang]



Context: compared to bottom-up expectations, a very complicated gauge theory with rich cosmological possibilities and concerns. (e.g. inflation, reheating, dark matter)

ML Application One:

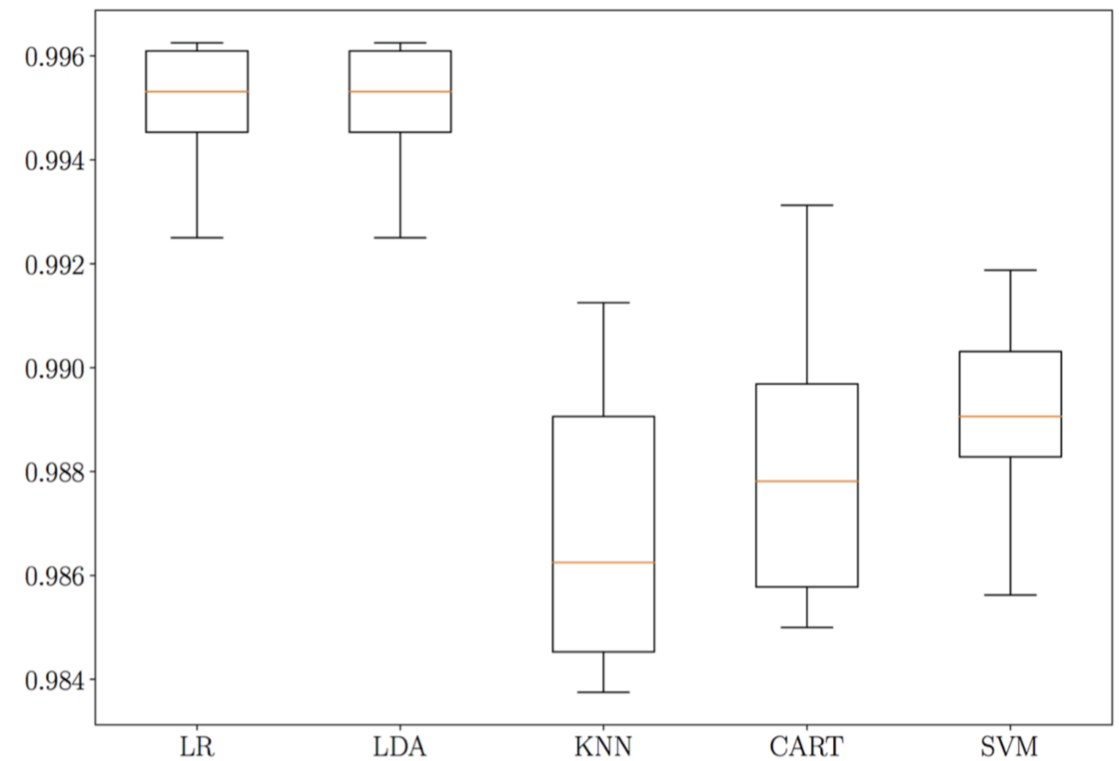
Q: Where is something like the SM?

Simple Supervised Learning
—> Generate Conjecture
—> Rigorous Theorem

1707.00655 with Carifio, Krioukov, Nelson

Interpretability + Theorems

- any GUTs or SM-like features?
  prob(E6 on special manifold) ~ 1/2000
- prepared training set, used simple supervised ML to learn when E6.
  (see right)
- some of the ML models showed "linchpin variable" that determined whether or not E6. Conjecture!
- Refine conjecture, get theorem.
- **Key idea:** simple ML to get rigorous results via conjecture generation.
  Simpler algs —> interpretability.



	LR	LDA	KNN	CART	SVM
50/50 Validation Set	.994	.994	.982	.987	.989
Unenriched Set	.988	.988	.981	.988	.983.

Theorem: Suppose that with high probability the group G on v_{E_6} is $G \in \{E_6, E_7, E_8\}$ and that E_6 may only arise with $\tilde{m} = (-2, 0, 0)$. Given these assumptions, there are three cases that determine whether or not G is E_6.

a) If $a_{max} \geq 5$, $\tilde{m}$ cannot exist in Δ_g and the group on v_{E_6} is above E_6.

b) Consider $a_{max} = 4$. Let $v_i = a_i v_{E_6} + b_i v_2 + c_i v_3$ be a leaf built above v_{E_6}, and $B = \tilde{m} \cdot v_2$ and $C = \tilde{m} \cdot v_3$. Then G is E_6 if and only if $(B, b_i) > 0$ or $(C, c_i) > 0 \; \forall i$. Depending on the case, G may or may not be E_6.

c) If $a_{max} \leq 3$, $\tilde{m} \in \Delta_g$ and the group is E_6.

ML Application Two:

Q: Where does this data set "fit"?

Relative to weak coupling landscape.

Intelligent exploration of weak coupling limits with asynchronous advantage actor-critic.

19xx with Long, Ruehle, Tian

The String Context

- 15 years ago: **weakly coupled** IIB.
  Now (and slightly after then), F-theory, how much broader?

- **Sharp Q:**
  how likely is existence of F —> IIB weak coupling limit?

- **Analytically proved** that P(weak coupling limit) $< 10^{-391}$.

- But likely way off, and also non-trivial lower bound?
  Also, what happens at the boundary, physically speaking?

"Connected" Anomaly Detection

- **Special data and a strategy?**

  Starting polytope (the "origin") has WCL, i.e. has anomaly.

- In general: extremely rare,

$$P(WCL) < 10^{-391}$$

  but connected to origin.

- Suggests moving out and penalizing when "out of bounds". (i.e. no WCL)

Common Situation:

Our Situation:

Reinforcement Learning

supervised ML **predicts**, RL **explores / searches**

in strings: see Halverson, Nelson, Ruehle to appear 1902. O(2000) improvement.

- an ***agent*** interacts in an ***environment***.
- it perceives a ***state*** from ***state space***.
- its ***policy*** picks and executes an action, given the state.
- agent arrives in new state, receives a ***reward.***
- succesive rewards accumulate into ***return.***
- return may penalize future rewards via ***discount factor***.
- policy optimized to maximize reward, i.e. **agent learns how to act!**

AlphaGo Zero

"Mastering the game of Go without human knowledge."

Silver et al. (Google DeepMind), Nature Oct. 2017.

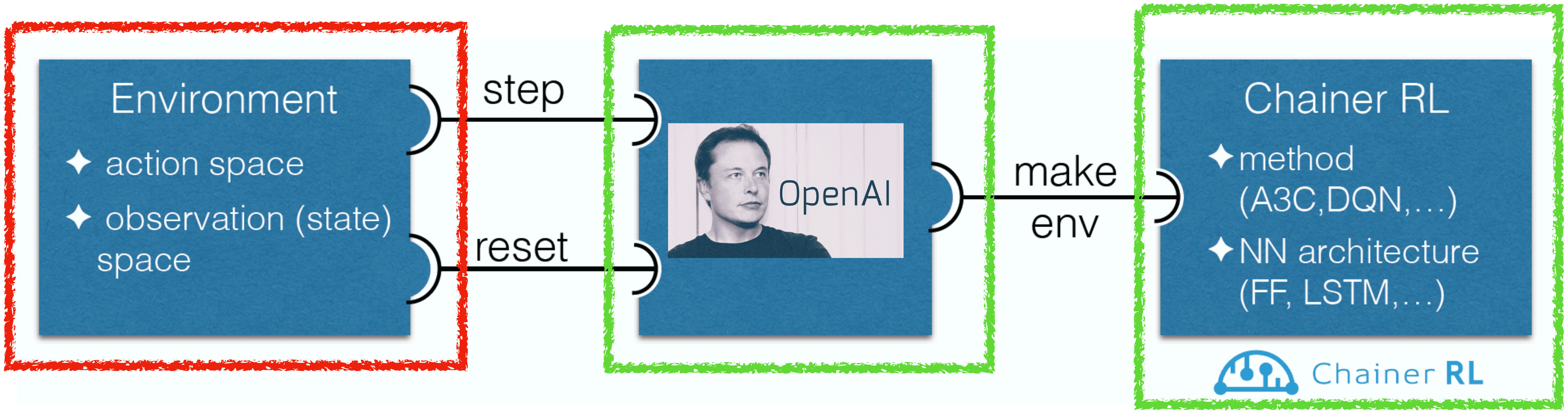
A long-standing goal of artificial intelligence is an algorithm that learns, tabula rasa, superhuman proficiency in challenging domains. Recently, AlphaGo became the first program to defeat a world champion in the game of Go. The tree search in AlphaGo evaluated positions and selected moves using deep neural networks. These neural networks were trained by supervised learning from human expert moves, and by reinforcement learning from self-play. **Here we introduce an algorithm based solely on reinforcement learning, without human data, guidance or domain knowledge beyond game rules.** AlphaGo becomes its own teacher: a neural network is trained to predict AlphaGo's own move selections and also the winner of AlphaGo's games. This neural network improves the strength of the tree search, resulting in higher quality move selection and stronger self-play in the next iteration. **Starting tabula rasa, our new program AlphaGo Zero achieved superhuman performance, winning 100–0 against the previously published, champion-defeating AlphaGo.**

Fact: Go has 10^{172} states, a "big" number, but for the task of playing excellently, superhuman progress achieved tabula rasa.

Implementation

model-free RL: want algorithms to work well regardless of environ.

three modules:
- Open AI defines what an environ is and how to interface.
- ChainerRL provides RL algorithms and NN architecture.
- **Physicists provide:** the environment. Rather simple!



algorithm: asynchronous advantage actor-critic (A3C) [Minh et al 2016]
(parallel CPU, not GPU)

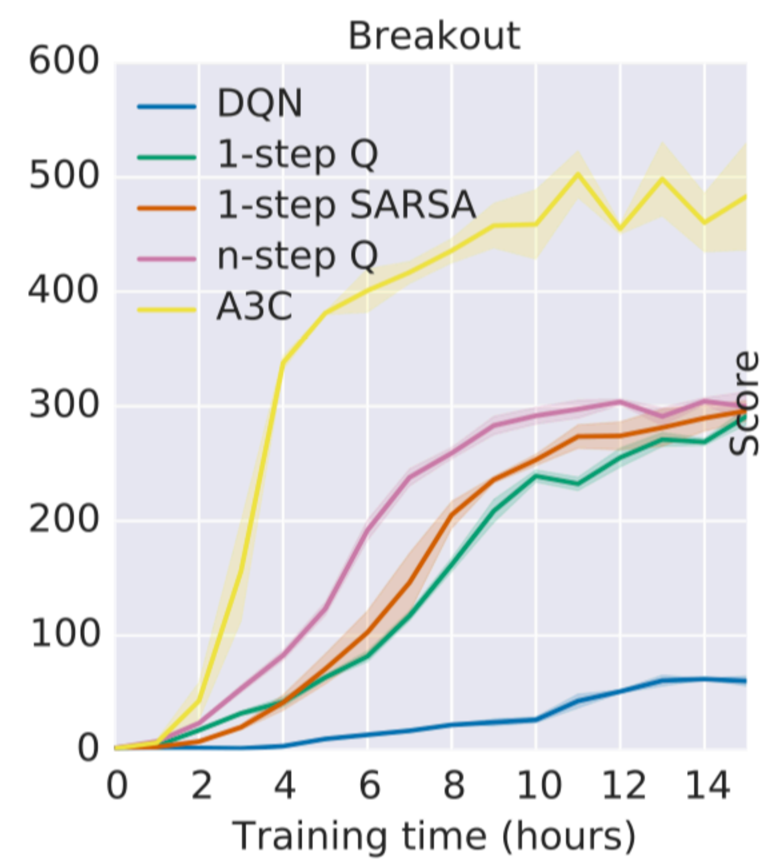
Asynchronous Advantage Actor-Critic (A3C)

[Mnih et al, DeepMind 2016]

*"Our parallel reinforcement learning paradigm also offers practical benefits. Whereas previous approaches to deep reinforcement learning rely heavily on specialized hardware such as GPUs (Mnih et al., 2015; Van Hasselt et al., 2015; Schaul et al., 2015) or massively distributed architectures (Nair et al., 2015), **our experiments run on a single machine with a standard multi-core CPU**. When applied to a variety of Atari 2600 domains, on many games asynchronous reinforcement learning achieves better results, in **far less time than previous GPU-based algorithms**, using far less resource than massively distributed approached" - Mnih et al,* Asynchronous Methods for Deep RL

- **Actor-Critic Methods:** NN for determining both policy (actor) and value (critic).
- **Asynchronous:** many worker bees explore, report back to king (critic) and queen (actor) bee.

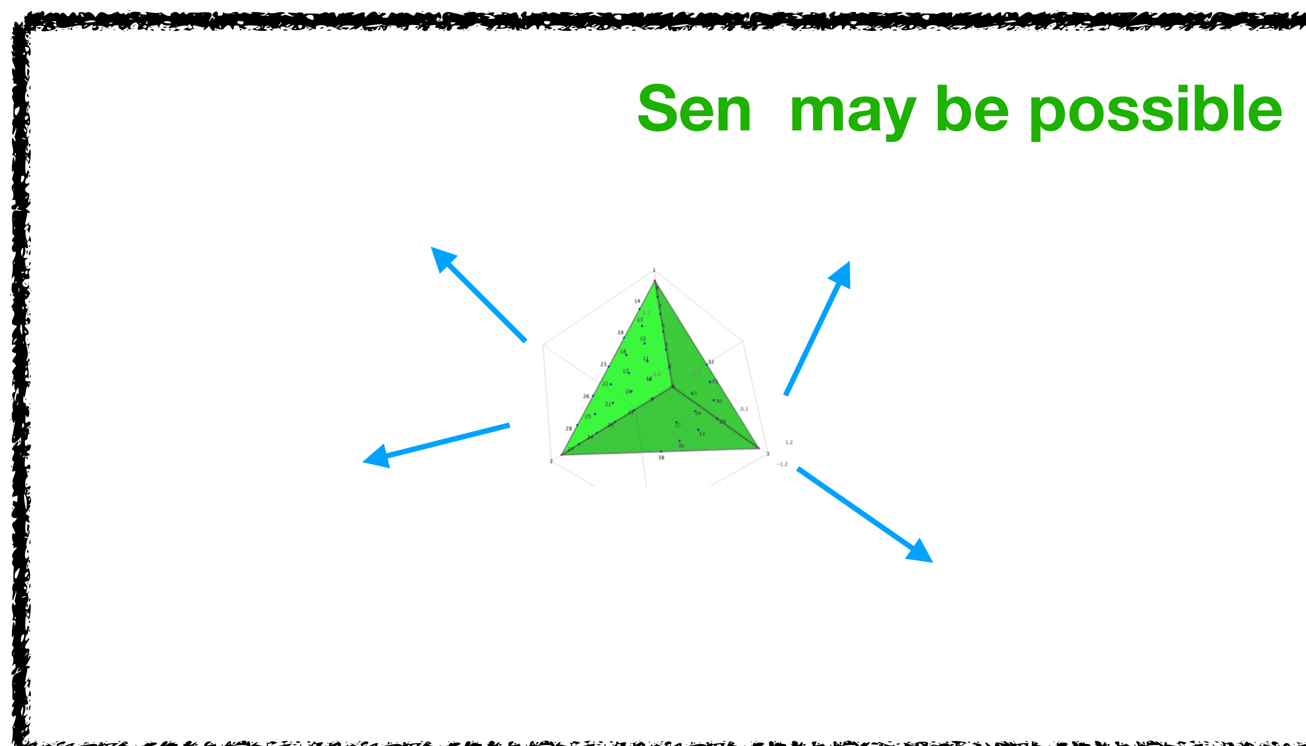
  i.e. use **communal knowledge.**
- **Advantage:** policy update depends on $A(s,a) = Q(s,a) - V(s)$
- > some 2016 GPU algs. Simple to run. Learns strategy



The Game

why might this work? WCL-possible geometries **connected subset** of our ensemble, can start at FRST of 3d refl. poly.

No Sen limit!



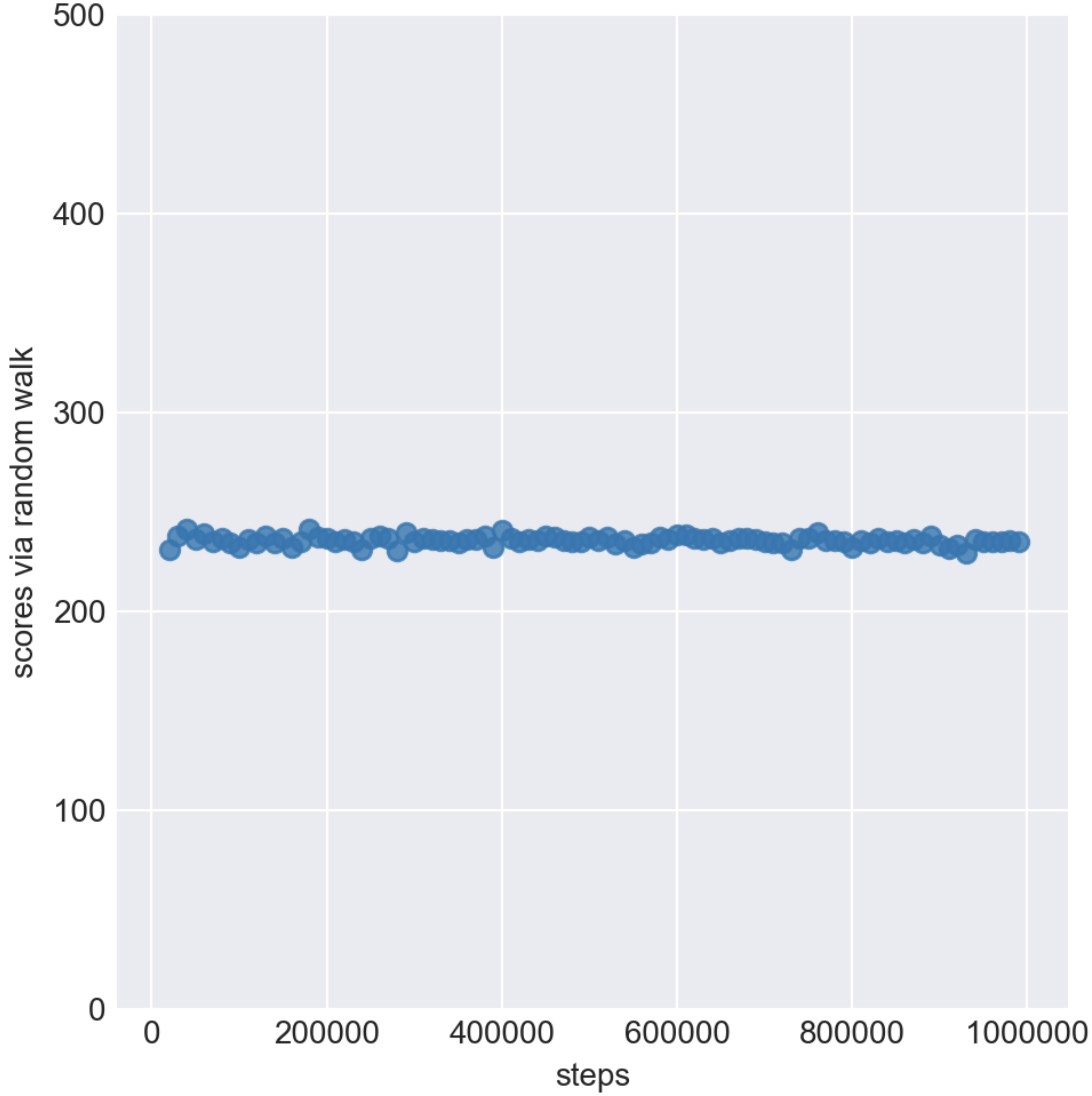
The Game:

move: seq. of transitions
goal: stay in bounds (WCL)
reward: 100 if in bounds
game over: out of bounds (no WCL)

want to show some **initial results.**

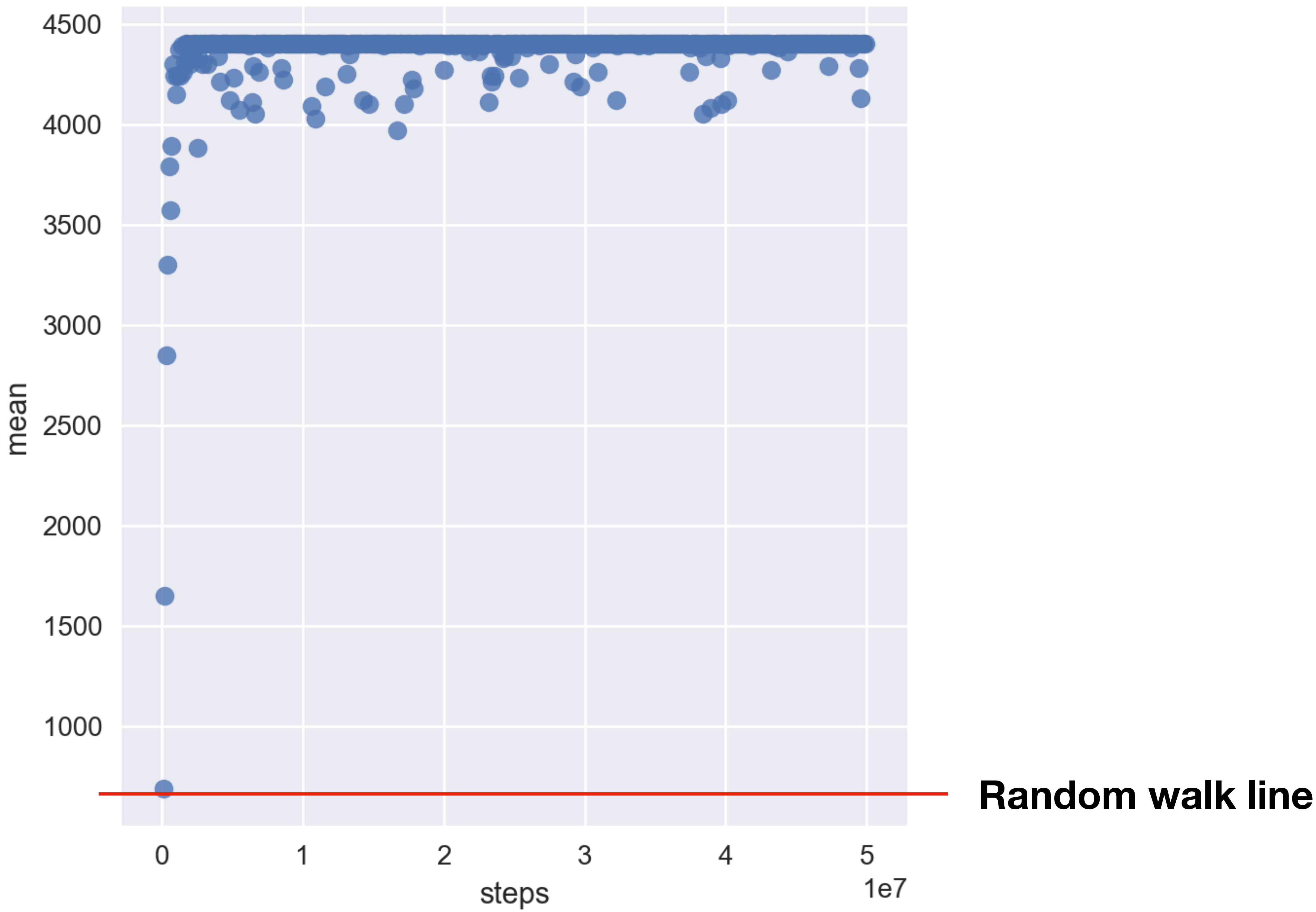
(presented in time series of results, to emphasize fun)

For Comparison: Random Walk



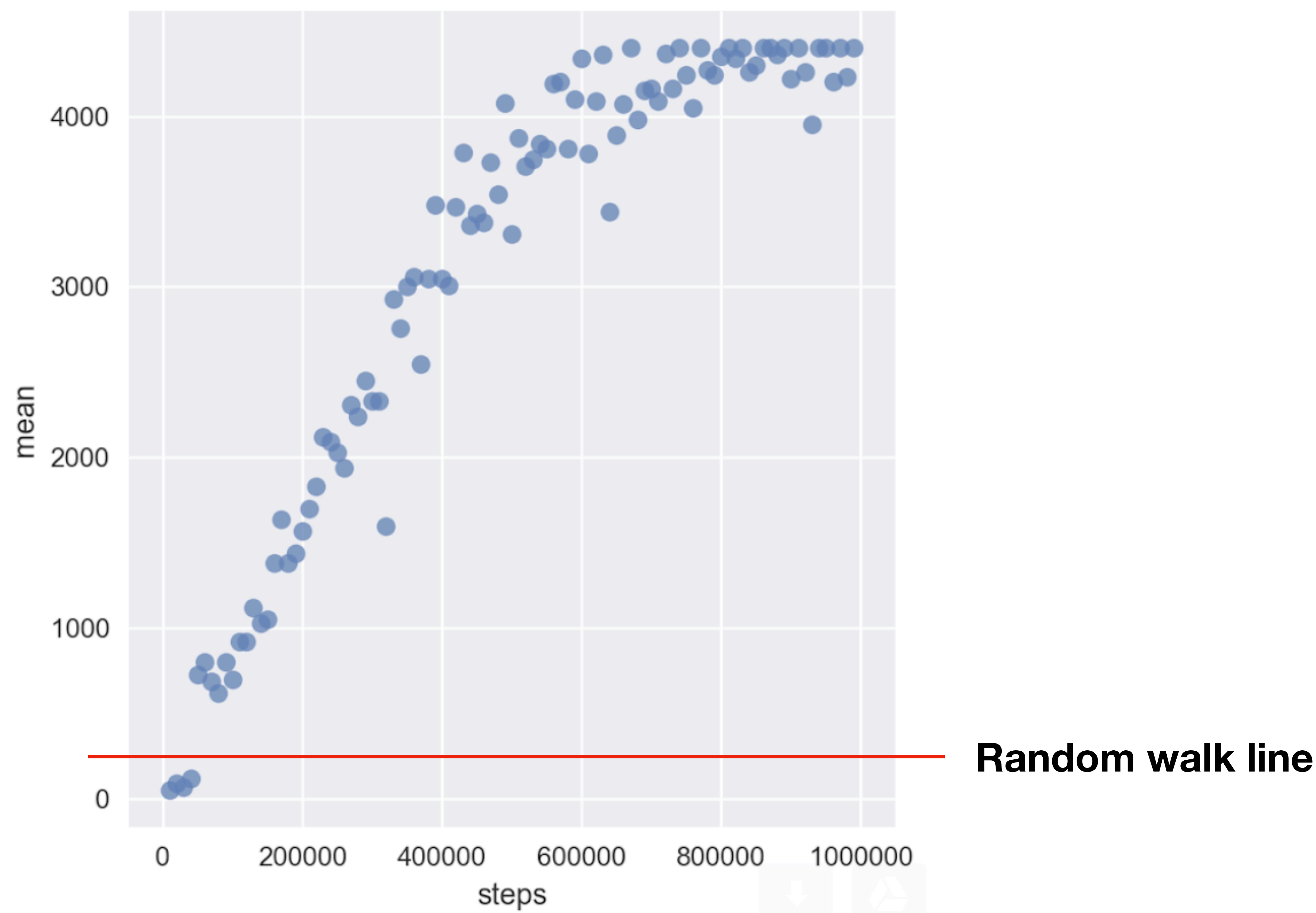
note scale: random walk takes 2-3 steps before No WCL

First Try: It Learns Quickly!



zoom in: decrease training time, increase eval interval

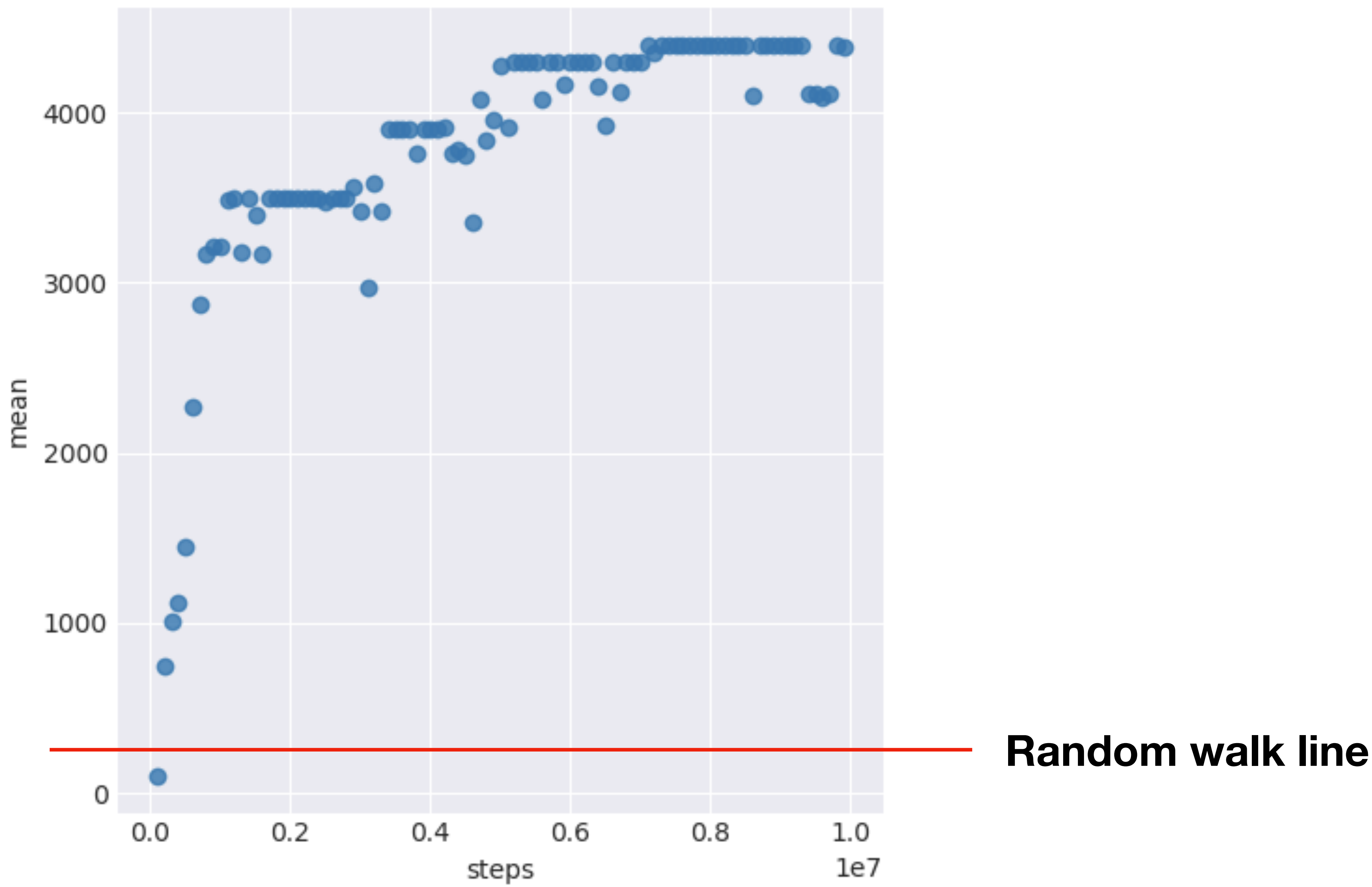
Second Try: See More Asymptote



much better, but can we tweak so it does better?

Third Try: Different Policy NN

Try LSTM instead of feedforward.

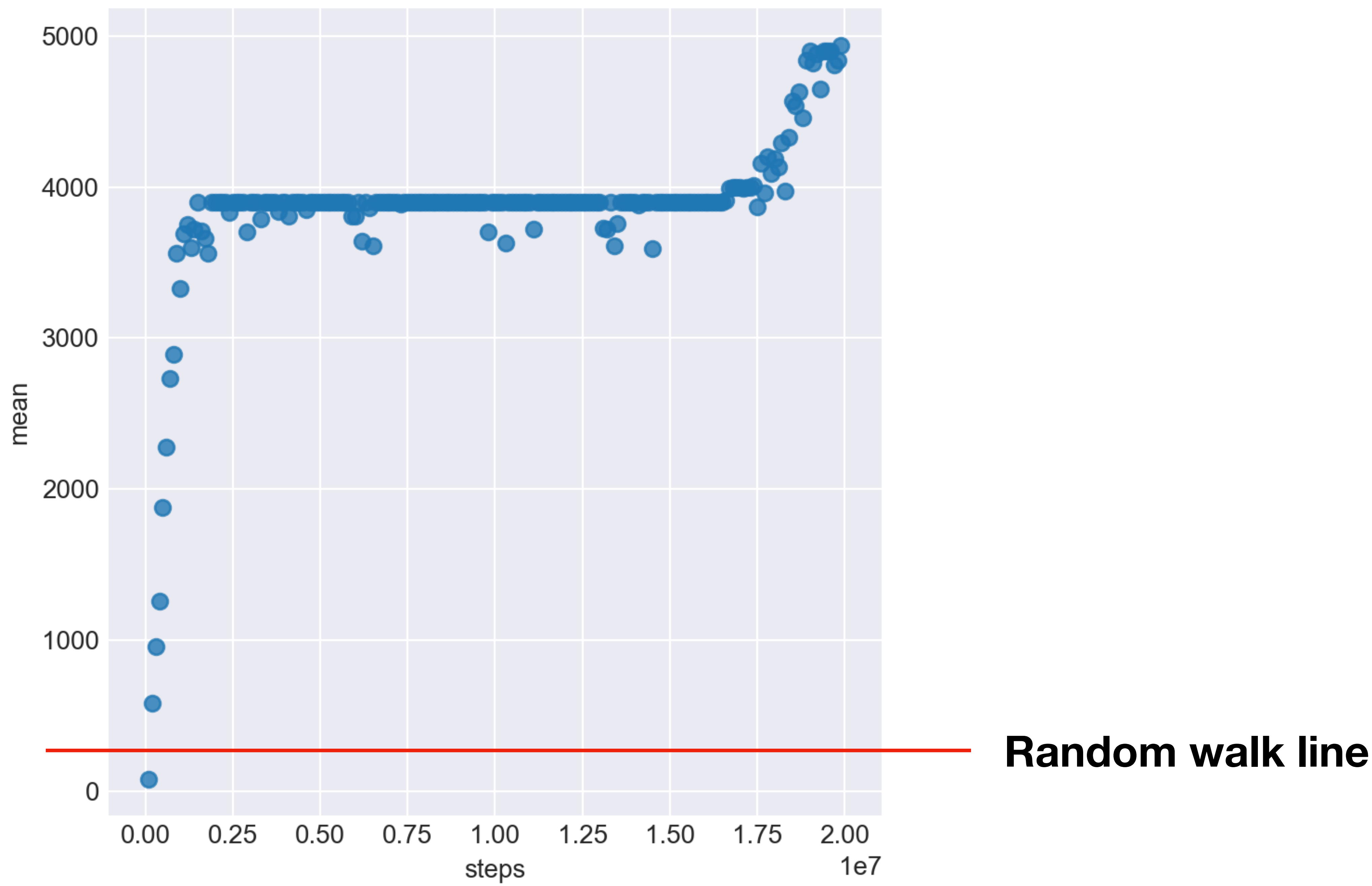


new feature: **four sharp plateaus**.

this is **punctuated equilibrium**.

Fourth Try: Don't Give Up

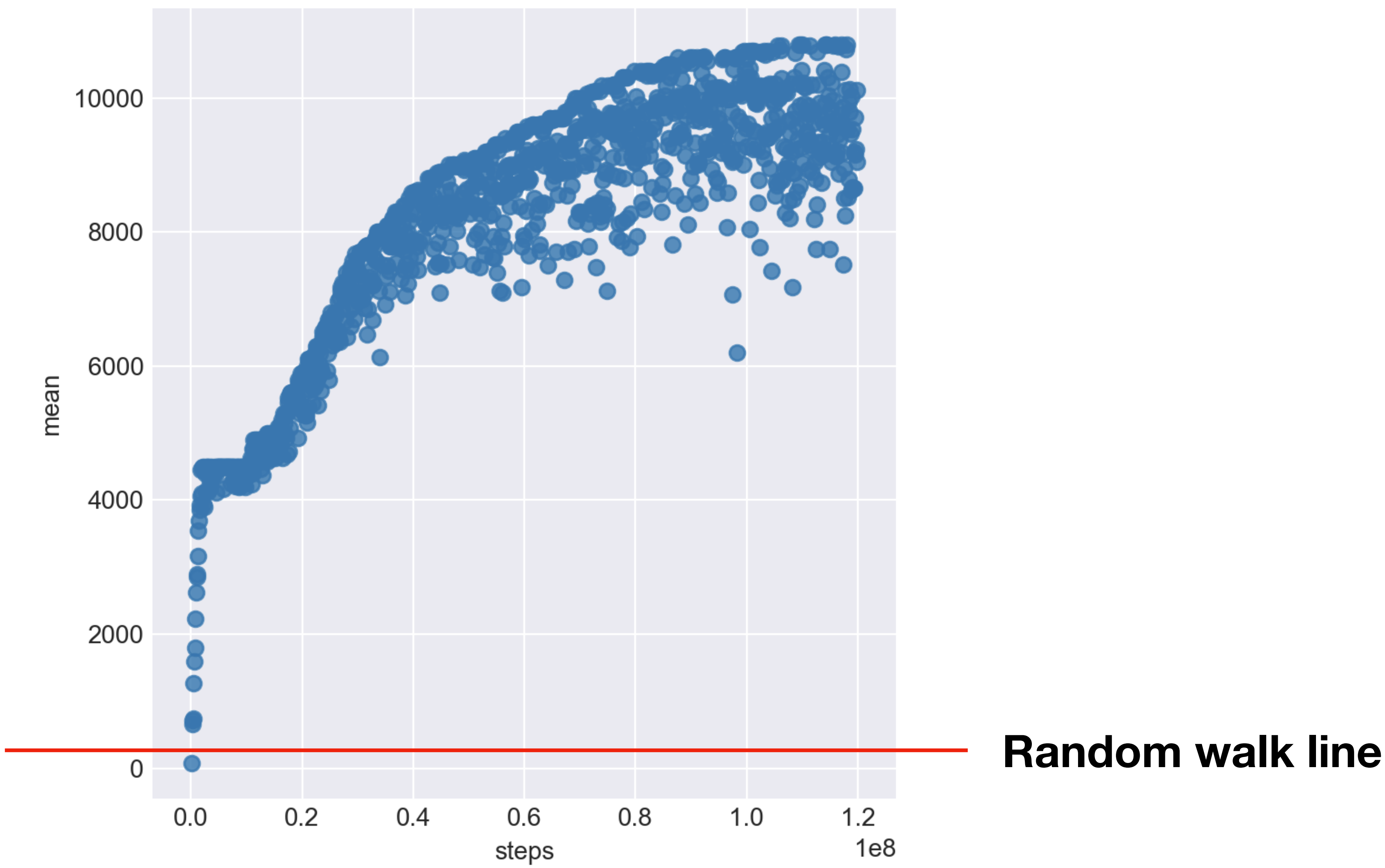
train a little longer, maybe it's got more juice in it.



work work work, keep on training.

Fifth Try: The Best Yet

and there's clearly still room to grow.



punchline: found deep weakly coupled crevice. Understand!

Upshot

- The 10,800 score solution sheds light on other WCL solutions. Gives lower bound below.

- **1)** weak coupling very rare:
  $10^{30} < N_{weak} < 10^{80}$ in 10^{755} ensemble

  2) typical weakly coupled model has at least 30 SO(8) seven-brane stacks that can typically be Higgsed in CS.

- Use explorer to generate WCL-NOWCL training pairs along largest known section of the boundary, study physics.

ML Application Three:

Q: Why are there SM particles?

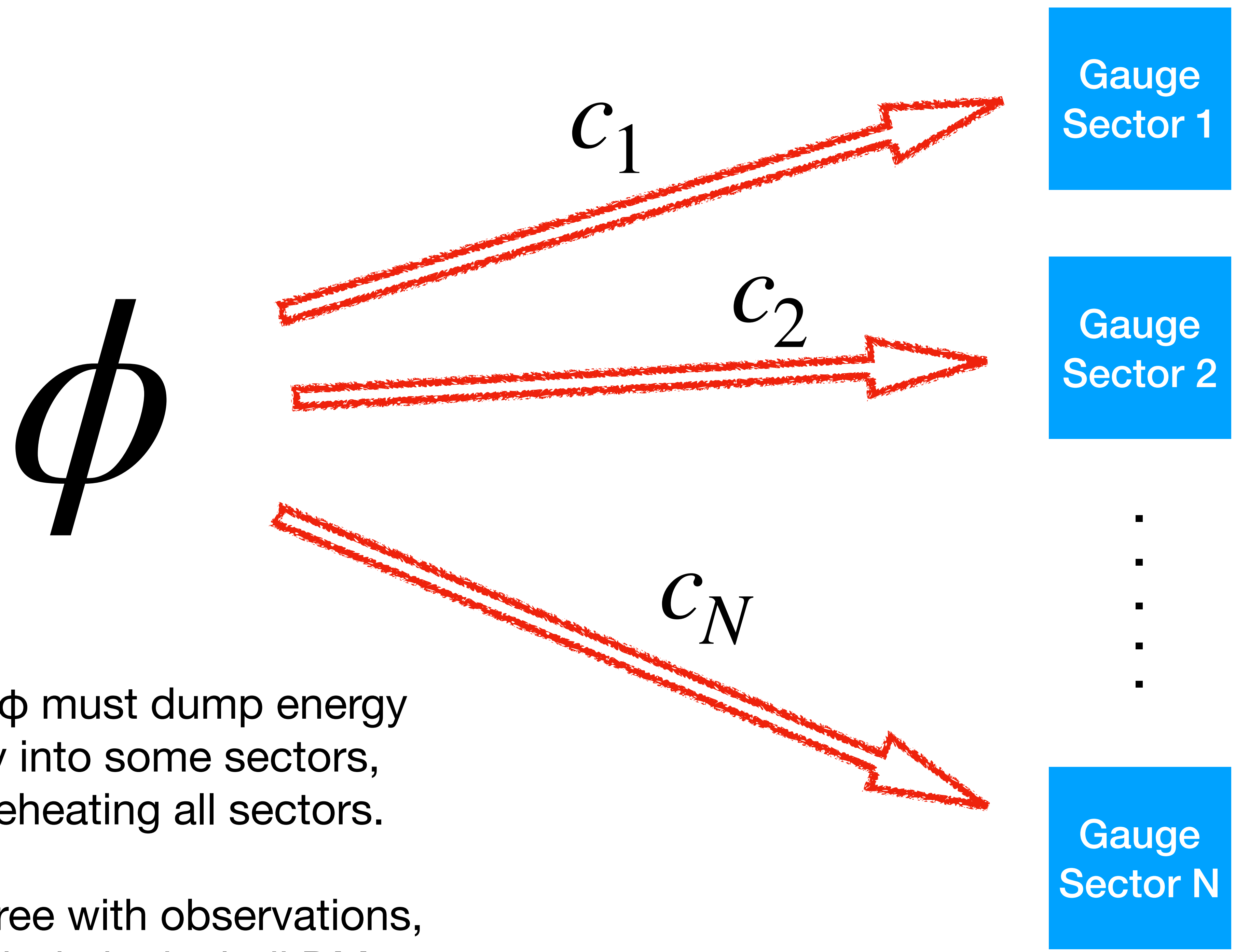
Answer: post-inflationary reheating, between "the end of the beginning" and "radiation."

Here: reheating in large ensemble.

NN —> surprising local understanding

1903 with Long, Nelson, Salinas

Reheating N Sectors



Constraint: ϕ must dump energy preferentially into some sectors, not equally reheating all sectors.

Or else: disagree with observations, e.g. too much dark glueball DM.

Axion-like Particle Effective Theory

$$\mathcal{L} = -\frac{1}{2}\delta_{ab}(\partial\phi^a)(\partial\phi^b) - \frac{1}{4}\sum_\alpha F_\alpha \wedge *F_\alpha - \sum_\alpha c_i^\alpha \phi^i F_\alpha \wedge F_\alpha$$

- Q: if axion inflaton couples to a gauge field,
  how much does it necessarily couple to other gauge fields?

- Order zero question: take $\phi_{\rm inf}$ to point along a single gauge group G_1 .
  How much does $\phi_{\rm inf}$ couple to G_n?

$$\mathcal{L}_{\rm int} = -c_{\rm inf}\,\phi_{\rm inf} F_1 \wedge F_1 - (c_2\phi_{\rm inf} + d_2^a\phi^a)F_2 \wedge F_2 - \ldots$$

- Physics goal: compute $\dfrac{c_i}{c_{\rm inf}}$. String EFT machine + topology does this.

Decision Variables

- Rather complicated map from geometry to reheat ratio.

- Q: is it determined in terms of simpler variables?

- Trained NN (Pytorch, feedforward, few layers)

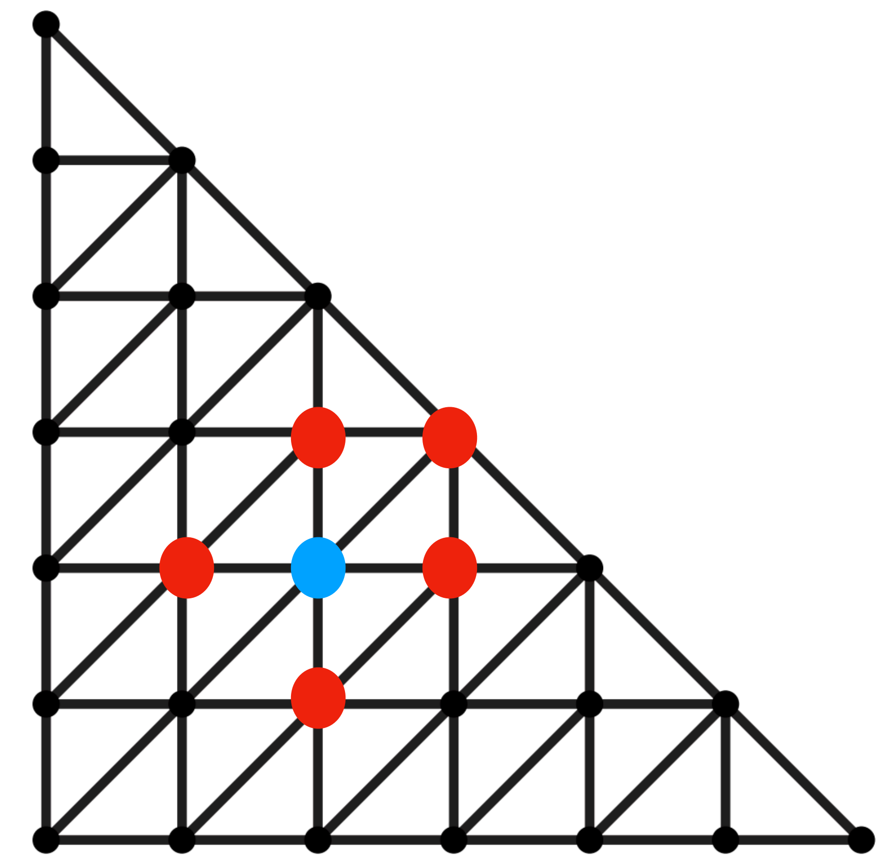
  on “tree heights”, gauge groups, graph distance various volumes including overall volume and a funny ratio we’ll see next slide, and more.

- Got great accuracy, so naively started column dropping.

- Found only two variables really matter!

Local Picture

$$\frac{c_o}{c_i} \simeq \frac{x_{io}}{x_{ii}} = \frac{\mathrm{vol}(D_i \cap D_o)}{\mathrm{vol}(D_i \cap D_i)} \times \frac{\mathrm{vol}(D_i)}{\mathrm{vol}(D_o)}$$



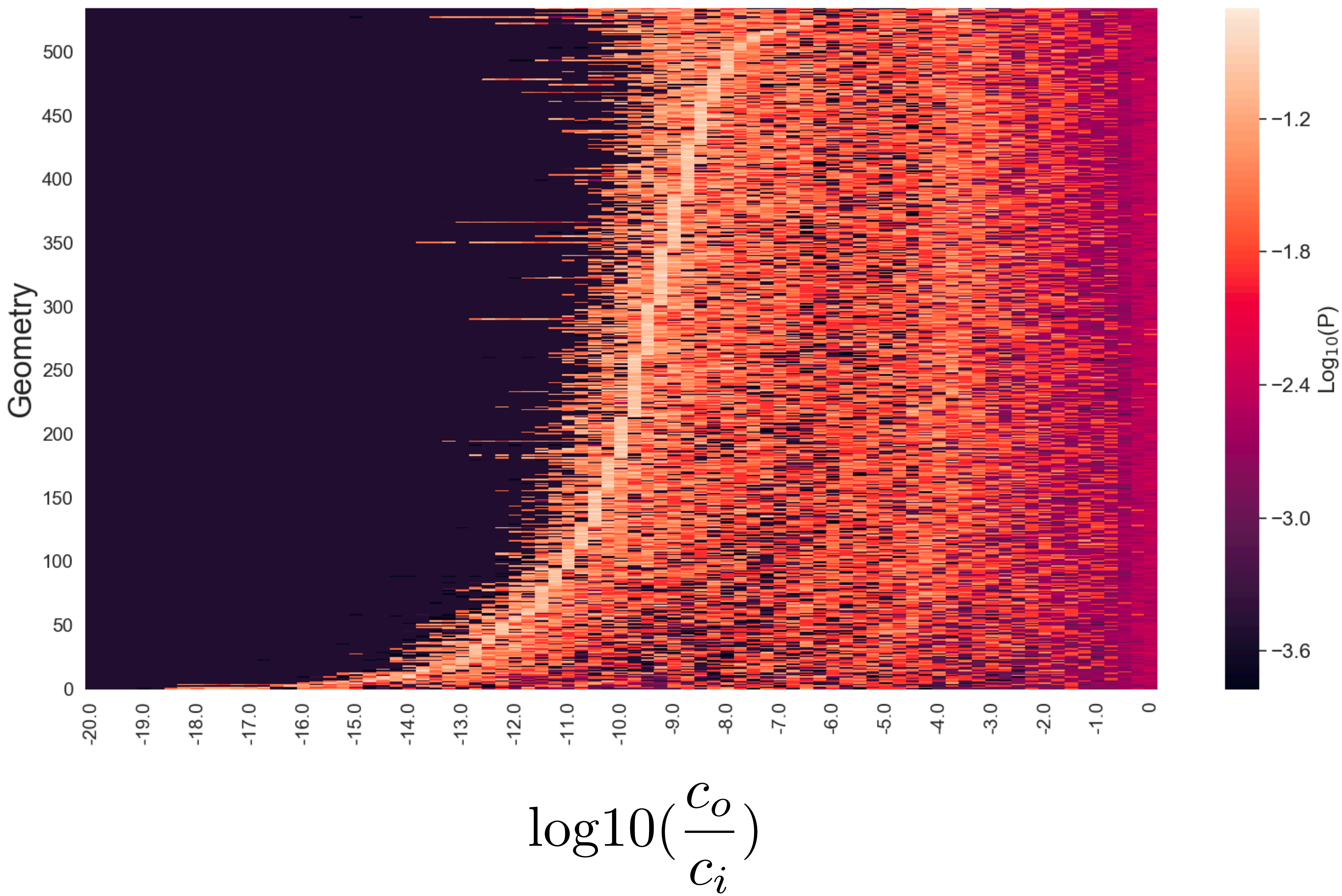
● = other gauge group reheated

● = inflaton gauge group

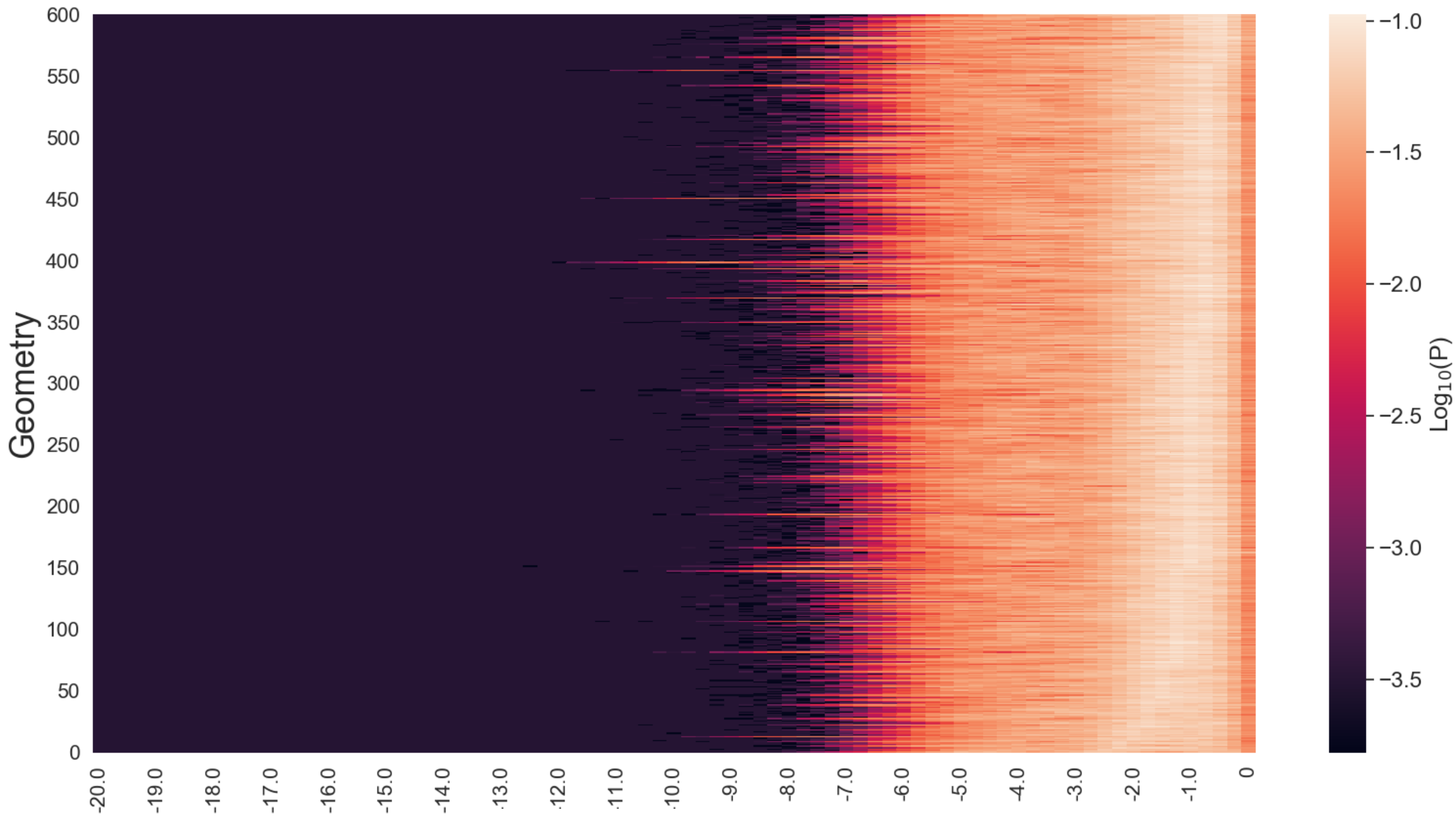
Result: for axion-like particles in gauge direction, couplings only occur for nearest neighbors, rest are very small.

Non-local physics is surprisingly negligible.

Reheating Distributions: Gauge Directions



Reheating Distributions: Random Directions



$$\log 10(\frac{c_o}{c_i})$$

Coefficients sampled from N(0, 1/sqrt(n))

Thank you!

Other string / HET topics investigated with ML, because I haven't done my friends' work justice.

- **Holographic QCD** and q-qbar potential. [Hashimoto, Sugishita, Tanaka, Tomiya] 1809.10536
- **Vacuum selection** in toy multiverse with network science. [Carifio, Cunningham, JH, Krioukov, Long, Nelson] 1711.06685
- **Islands of SMs** with autoencoders [Mutter, Parr, Vaudrevange] 1811.05993
- **Small cosmological constants** with RL and complexity. [JH, Long, Ruehle] in progress.
- **Persistent homology & vacua** [Cole, Shiu] 1812.06960
- **Line bundle cohomology** [Ruehle] 1706.07024 [Klaewer, Schlechter] 1809.02547
- **Branes w/ Branes**: RL for consistency + particle optimization [JH, Nelson, Ruehle], in progress
- **Upper bound for certain CY3s** [Altman, Carifio, JH, Nelson]
- **Model building with GANs** [Erbin, Krippendorf]

and many more . . .