MACHINE LEARNING AS A DISCOVERY TOOL IN HEP-TH

Vishnu Jejjala
University of the Witwatersrand

7 October 2020
Physics ∩ ML

MACHINE LEARNING ~~AS~~ IS A DISCOVERY TOOL IN HEP-TH

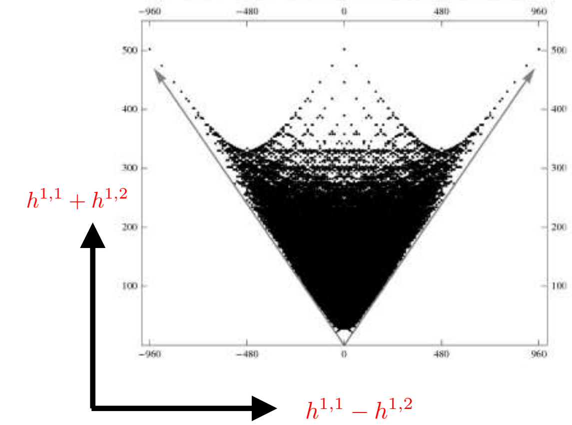
Vishnu Jejjala
University of the Witwatersrand

7 October 2020
Physics ∩ ML

Mathematical Phenomenology

- Note patterns then look for an explanation

  — Mirror symmetry is prototype example



- Use machine learning to train a computer to calculate in **`hep-th`**

  — Black box gives **probably approximately correct** answers

  — Filter interesting cases to investigate further

  — Start to dissect how it all works

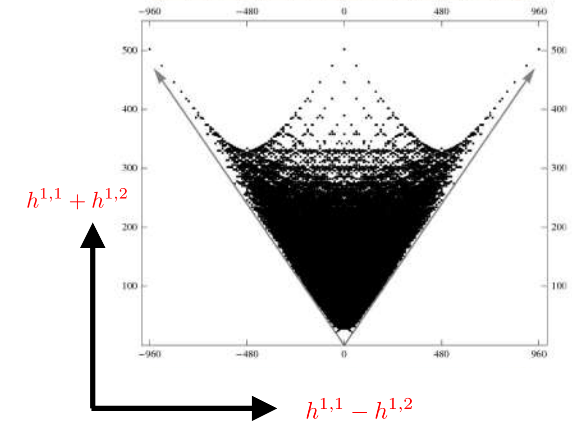
  — Want to bridge this success to new analytic results and methods

- No deep theoretical understanding for why machine learning is effective, for which problems it is useful, etc.

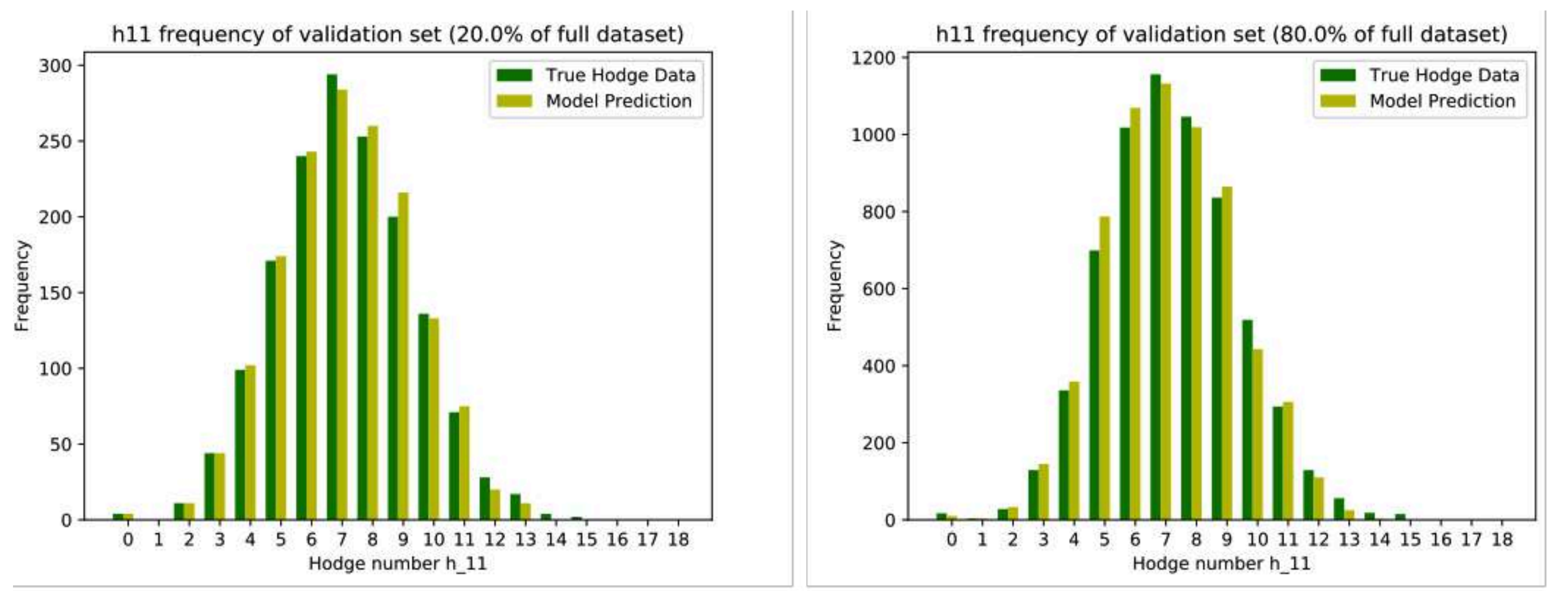
Mathematical Phenomenology

- Note patterns then look for an explanation

  — Mirror symmetry is prototype example



- Learn Hodge numbers of CICYs from configuration matrices



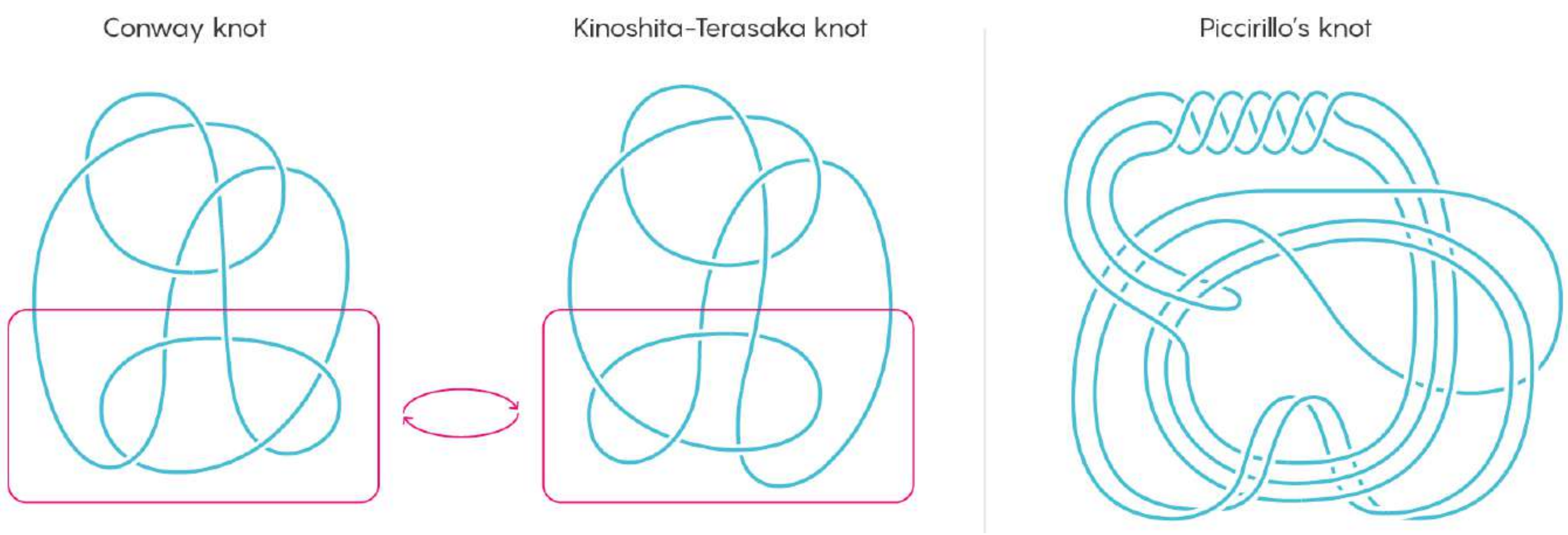
  — Polynomial rather than doubly exponential in time

  — Better ways to calculate in algebraic geometry

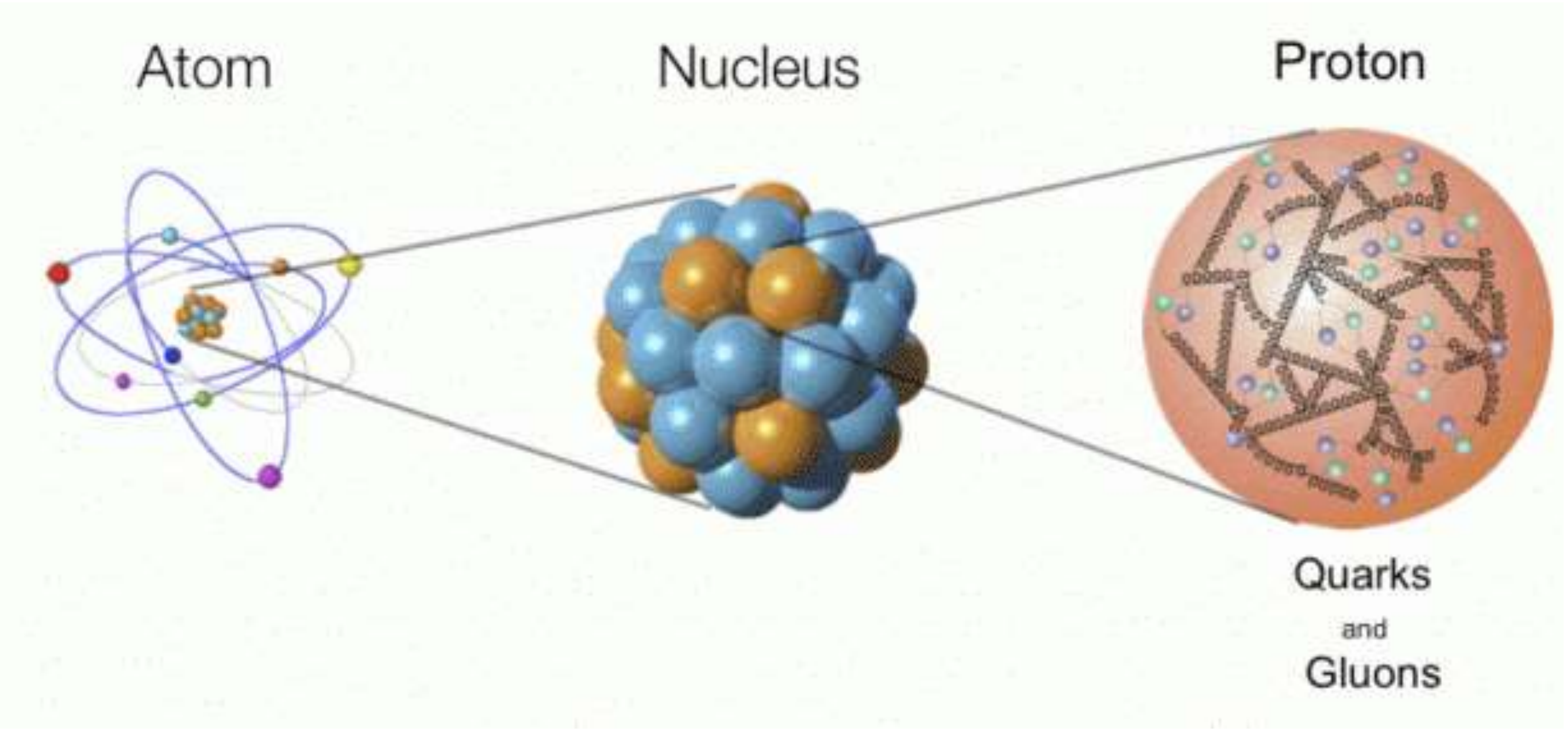
Bull, He, VJ, Mishra (2018)
Bull, He, VJ, Mishra (2019)

Two Case Studies

- Knots



- QCD

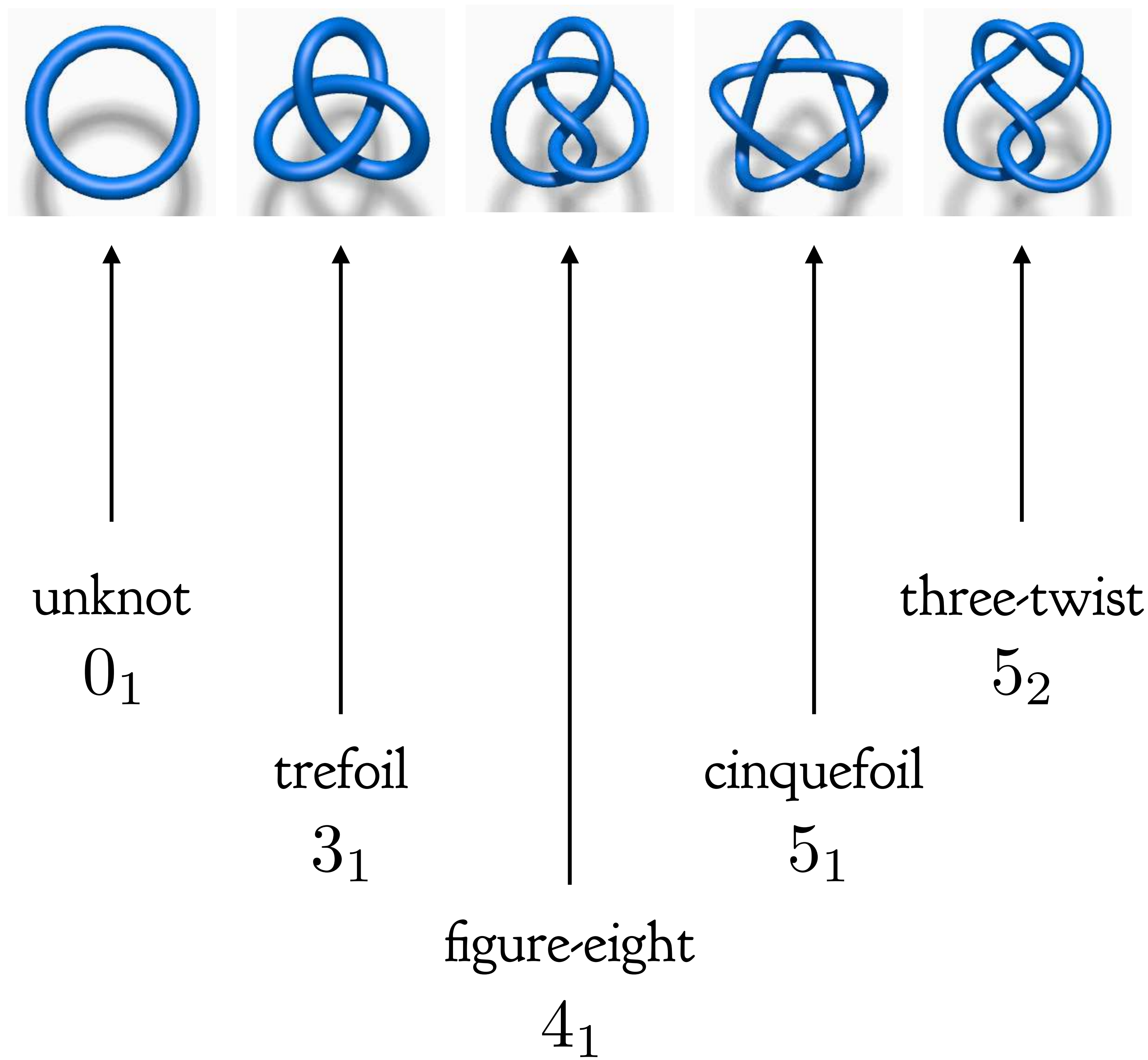


Please ask questions!

FROM JONES POLYNOMIALS TO HYPERBOLIC VOLUMES

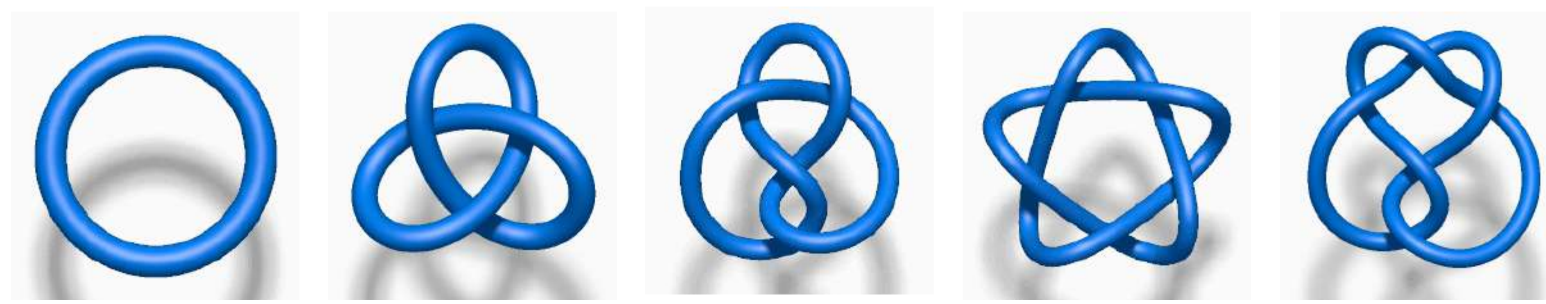
Dramatis Personae

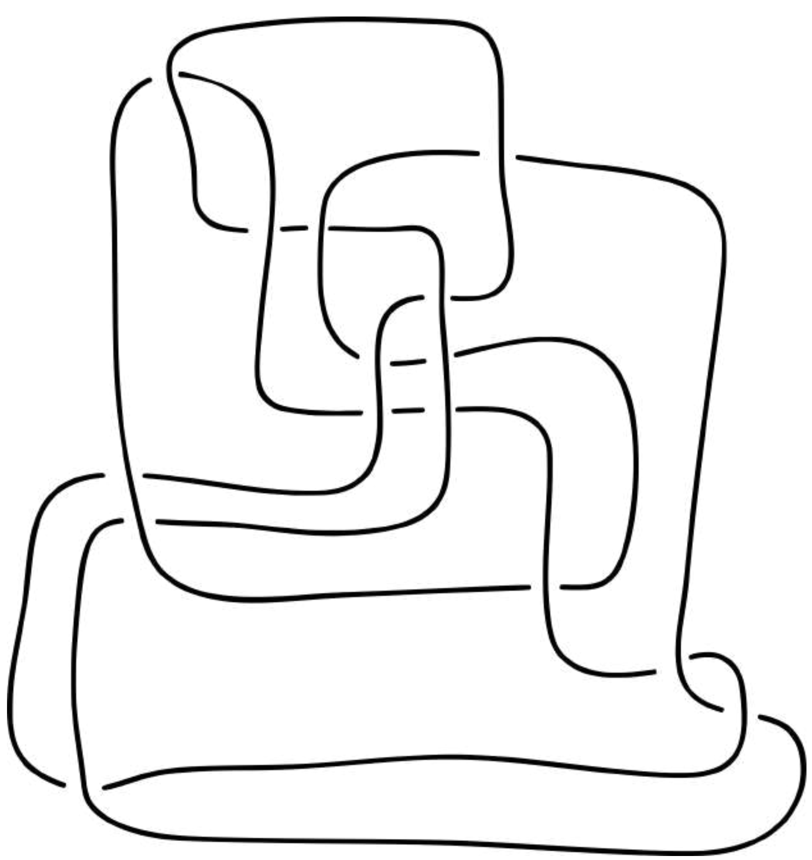
Knot: $S^1 \subset S^3$; e.g.,



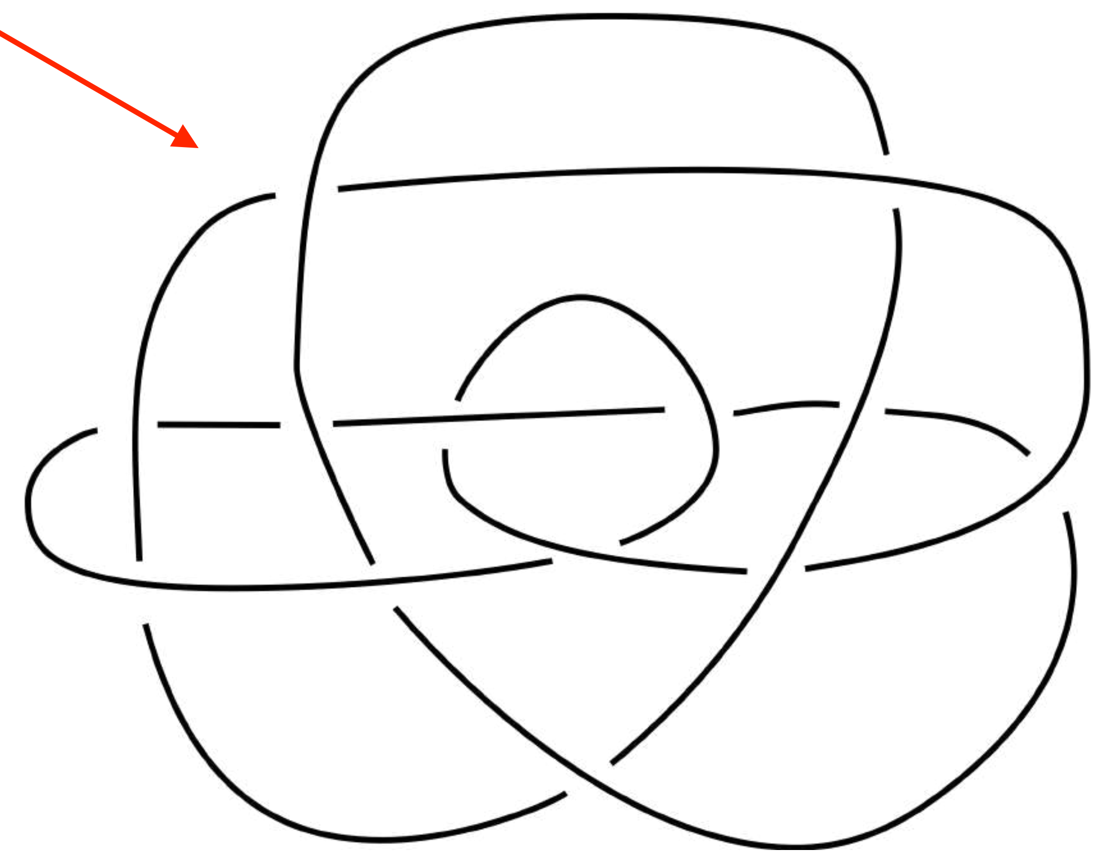
Dramatis Personae

Knot: $S^1 \subset S^3$; e.g.,





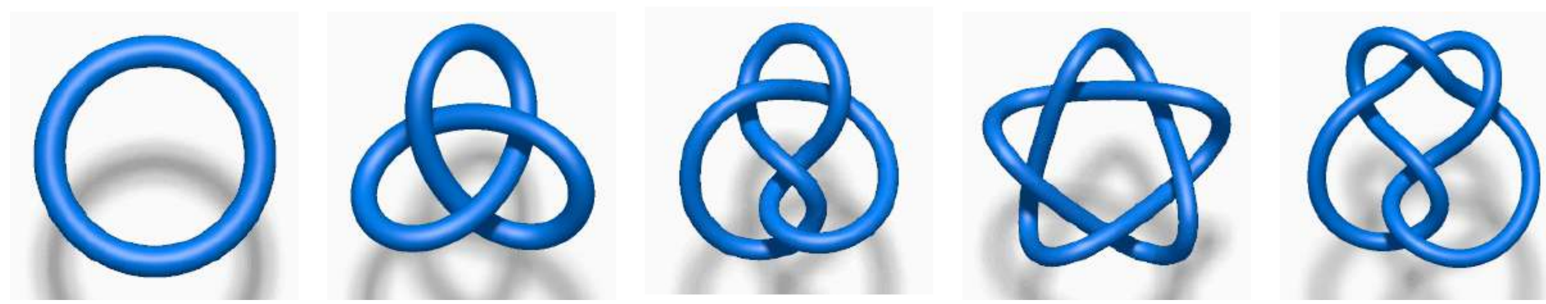
Thistlethwaite unknot



Ochiai unknot

Dramatis Personae

Knot: $S^1 \subset S^3$; e.g.,



Jones polynomial: $J_K(q) = (-q^{\frac{3}{4}})^{w(K)} \dfrac{\langle K \rangle}{\langle \bigcirc \rangle}$

$$\langle \times \rangle = q^{\frac{1}{4}} \langle \asymp \rangle + \frac{1}{q^{\frac{1}{4}}} \langle)(\rangle$$

$w(K) = \text{overhand} - \text{underhand}$

Jones (1985)

topological invariant: independent of how the knot is drawn

Question: how to calculate these?

Answer: quantum field theory!

Chern–Simons Theory

- What is the simplest non-trivial quantum field theory?

  — Chern–Simons theory in three dimensions

- Focus on **topology** instead of geometry

genus 0

genus 1

Topological Invariants

- On a manifold $\mathcal{M}$ with metric $g_{\mu\nu}$, a topological invariant enjoys:

$$\frac{\delta}{\delta g_{\mu\nu}} \langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle = 0$$

- In Chern–Simons theory, the operators are Wilson loops

$$W_R^\gamma = \mathrm{Tr}_R\, U_\gamma \ , \quad U_\gamma = \mathcal{P} \exp i \oint_\gamma A$$

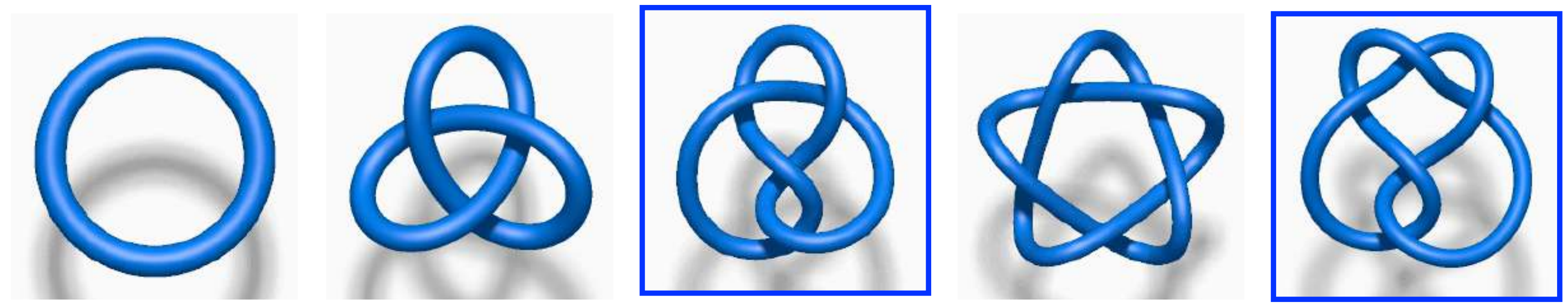
- The knot invariants are then

$$\langle W_{R_1}^{K_1} \dots W_{R_n}^{K_n} \rangle = \frac{1}{Z(\mathcal{M})} \int [DA]\, W_{R_1}^{K_1} \dots W_{R_n}^{K_n} e^{iS_{\mathrm{CS}}}$$

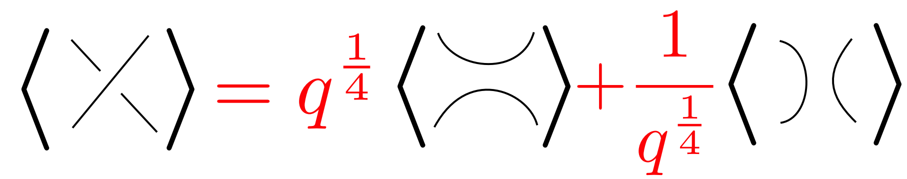
$$S_{\mathrm{CS}} = \frac{k}{4\pi} \int_{\mathcal{M}} \mathrm{Tr} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right) \ , \quad Z(\mathcal{M}) = \int [DA]\, e^{iS_{\mathrm{CS}}} \ , \quad k \in \mathbb{Z}$$

Dramatis Personae

Knot: $S^1 \subset S^3$; e.g.,



Jones polynomial: $J_K(q) = (-q^{\frac{3}{4}})^{w(K)} \frac{\langle K \rangle}{\langle \bigcirc \rangle}$



$w(K) =$ overhand − underhand

vev of Wilson loop operator along K in
$\square$ for $SU(2)$ Chern–Simons on S^3

Jones (1985)
Witten (1989)

$$J_{4_1}(q) = q^{-2} - q^{-1} + 1 - q + q^2 \,, \quad q = e^{\frac{2\pi i}{k+2}}$$

Hyperbolic volume: volume of $S^3 - K$ is another knot invariant

computed from tetrahedral decomposition of knot complement

Thurston (1978)

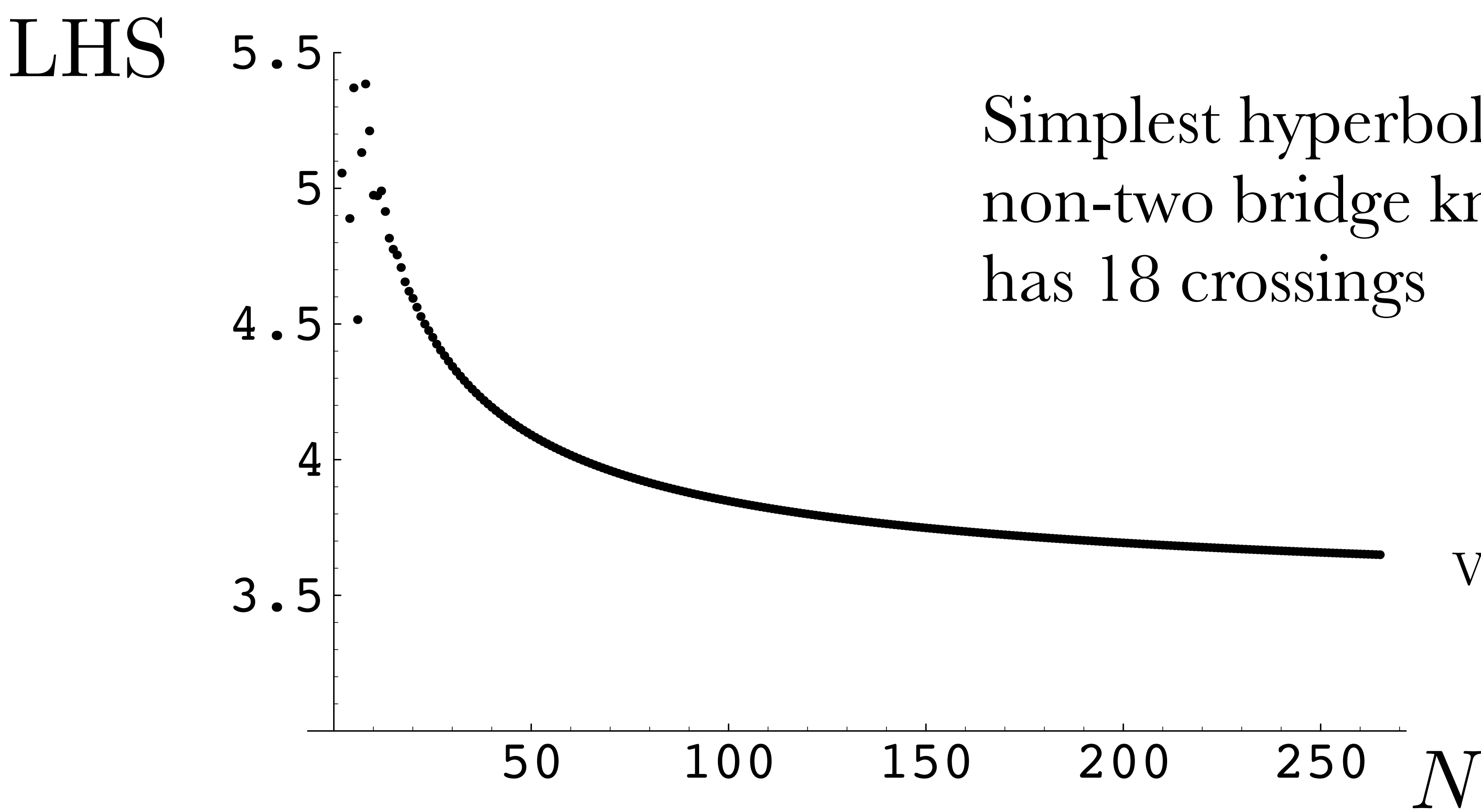
Dramatis Personae

Colored Jones polynomial: $J_{K,2} = J_K(q)$ Witten (1989)

$$\lim_{N\to\infty} \frac{2\pi \log |J_{K,N}(e^{\frac{2\pi i}{N}})|}{N} = \mathrm{Vol}(K)$$

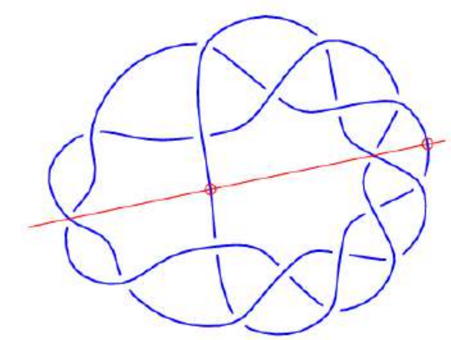
Kashaev (1997)
Murakami x 2 (2001)
Gukov (2005)

[Volume conjecture]



$\mathrm{Vol}(K_0) = 3.474247\dots$

Simplest hyperbolic non-two bridge knot, has 18 crossings



$\sigma_1^2(\sigma_1\sigma_2)^8$

Garoufalidis, Lan (2004)

Dramatis Personae

Colored Jones polynomial: $J_{K,2} = J_K(q)$ Witten (1989)

$$\lim_{N\to\infty} \frac{2\pi \log |J_{K,N}(e^{\frac{2\pi i}{N}})|}{N} = \mathrm{Vol}(K)$$

Kashaev (1997)
Murakami x 2 (2001)
Gukov (2005)

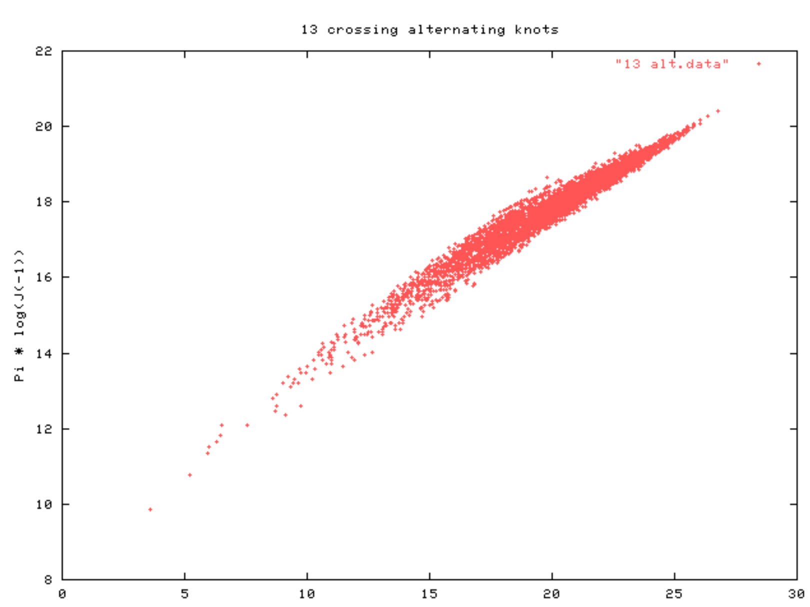
[Volume conjecture]

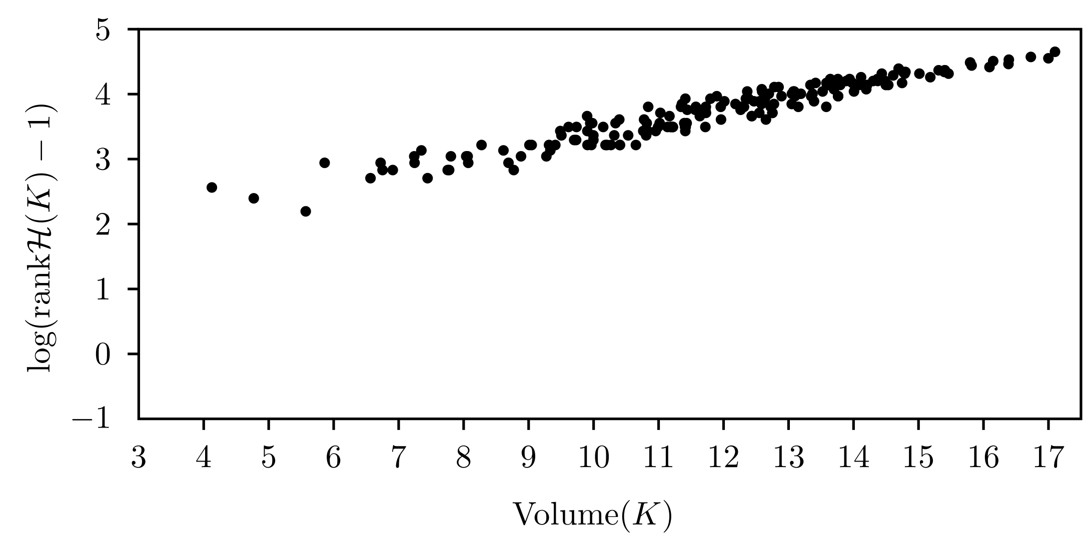
Khovanov homology: a homology theory $\mathcal{H}_K$ whose graded Euler characteristic is $J_K(q)$; explains why coefficients in $J_K(q)$ are integers

Khovanov (2000)

$$\log J_K(-1)\,,\ \log(\mathrm{rank}(\mathcal{H}_K) - 1) \propto \mathrm{Vol}(K)$$

Dunfield (2000)
Khovanov (2002)





Dramatis Personae

Colored Jones polynomial: $J_{K,2} = J_K(q)$ Witten (1989)

$$\lim_{N\to\infty} \frac{2\pi \log |J_{K,N}(e^{\frac{2\pi i}{N}})|}{N} = \mathrm{Vol}(K)$$

Kashaev (1997)
Murakami x 2 (2001)
Gukov (2005)

[Volume conjecture]

Khovanov homology: a homology theory $\mathcal{H}_K$ whose graded Euler characteristic is $J_K(q)$; explains why coefficients in $J_K(q)$ are integers

Khovanov (2000)

$$\log J_K(-1)\,,\ \log(\mathrm{rank}(\mathcal{H}_K)-1) \propto \mathrm{Vol}(K)$$

Dunfield (2000)
Khovanov (2002)

Volume-ish theorem: $J_K(q) = a_n q^n + \ldots + a_m q^m$

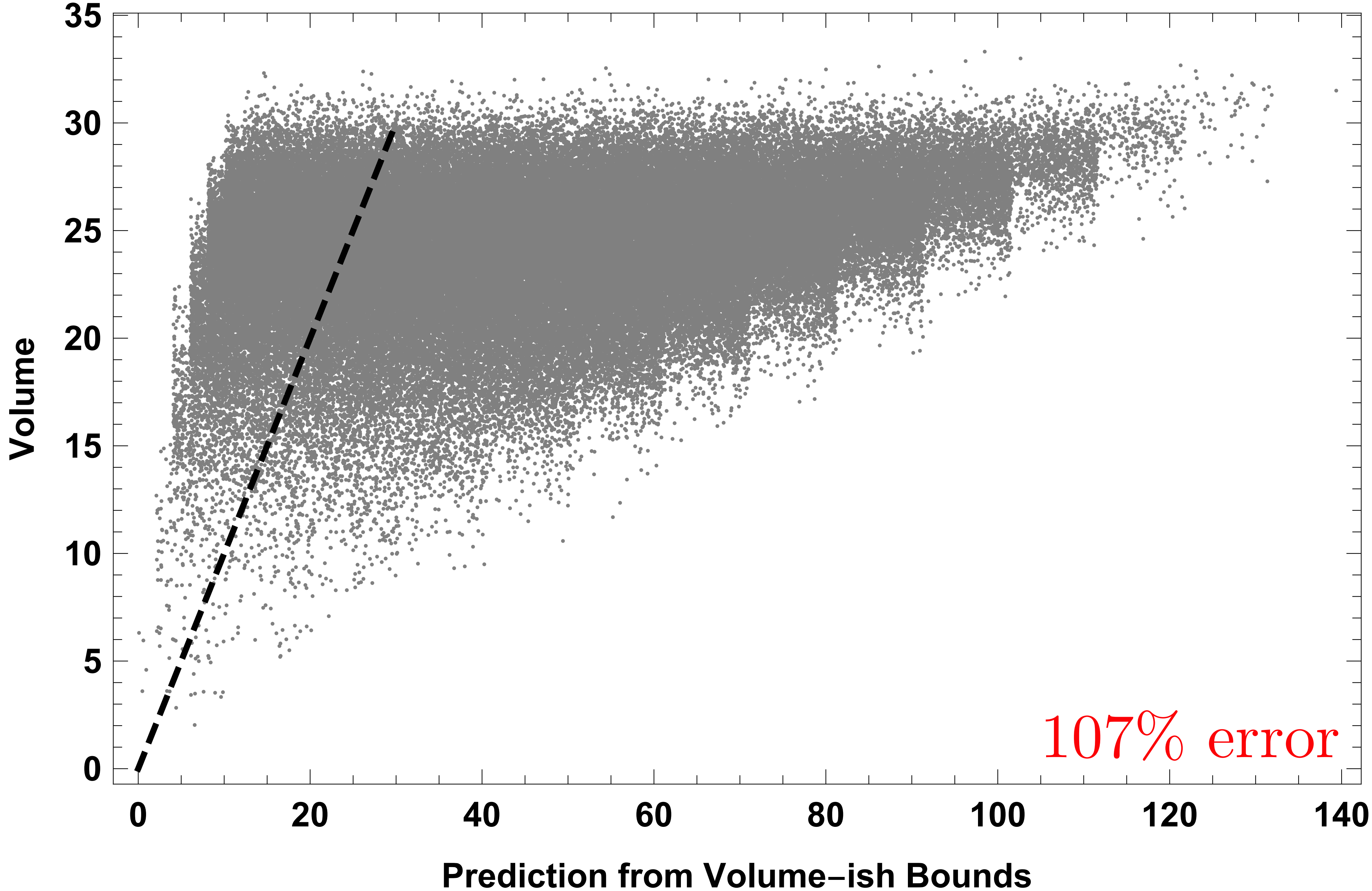
$$2v_0(\max(|a_{m-1}|, |a_{n+1}|)-1) \le \mathrm{Vol}(K) \le 10v_0(|a_{m-1}|+|a_{n+1}|-1)$$

for alternating knots; $v_0 = -\int_0^{\pi/3} dt\ \log(2\sin\frac{t}{2})$, volume of regular ideal tetrahedron

≈ 1.0149416

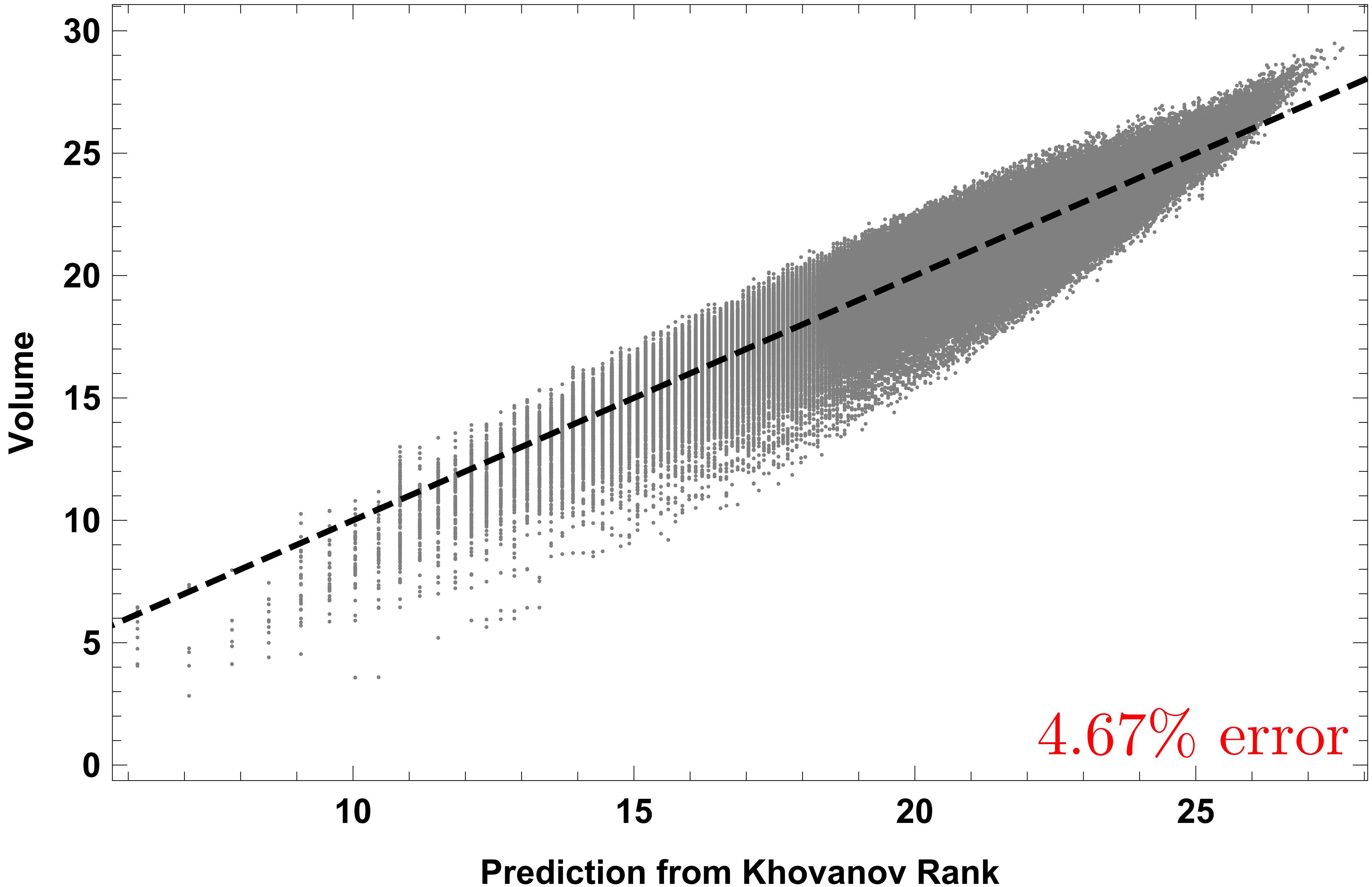
Dasbach, Lin (2007)

Volume-ish Bound



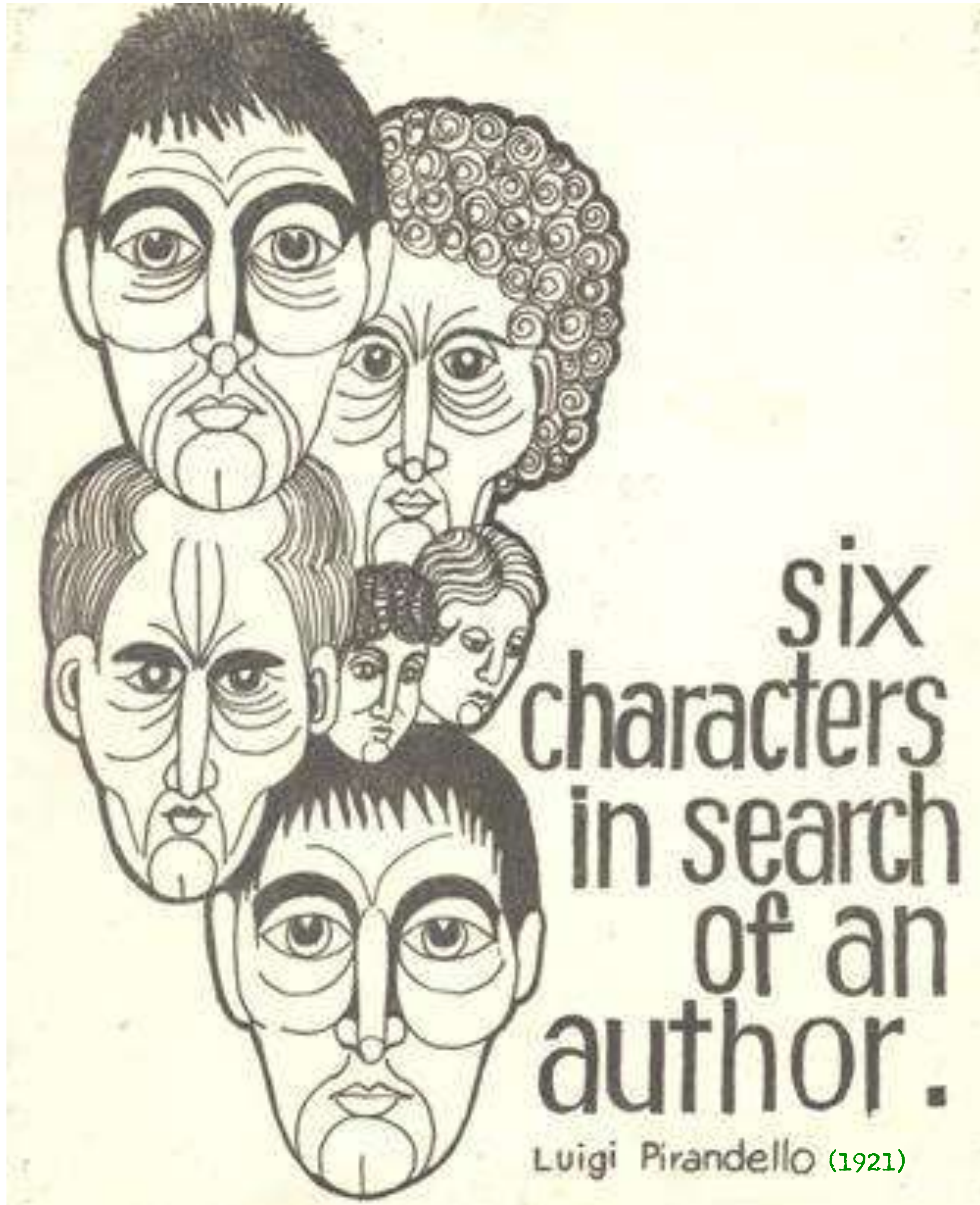
$111,521$ alternating knots (35.6%)

Khovanov Homology

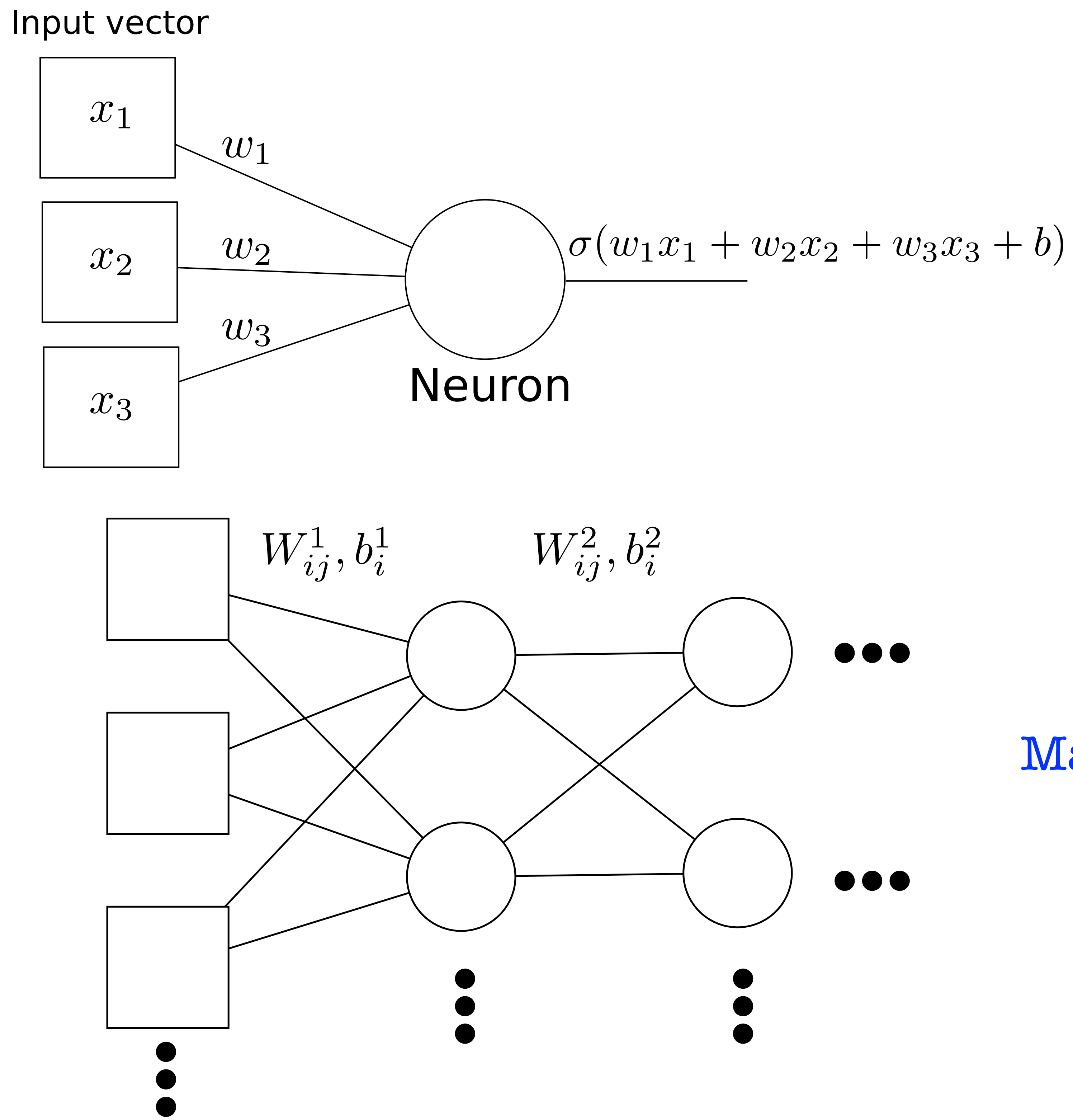


$196,011$ knots (62.6%)

Dramatis Personae



Feedforward Neural Networks



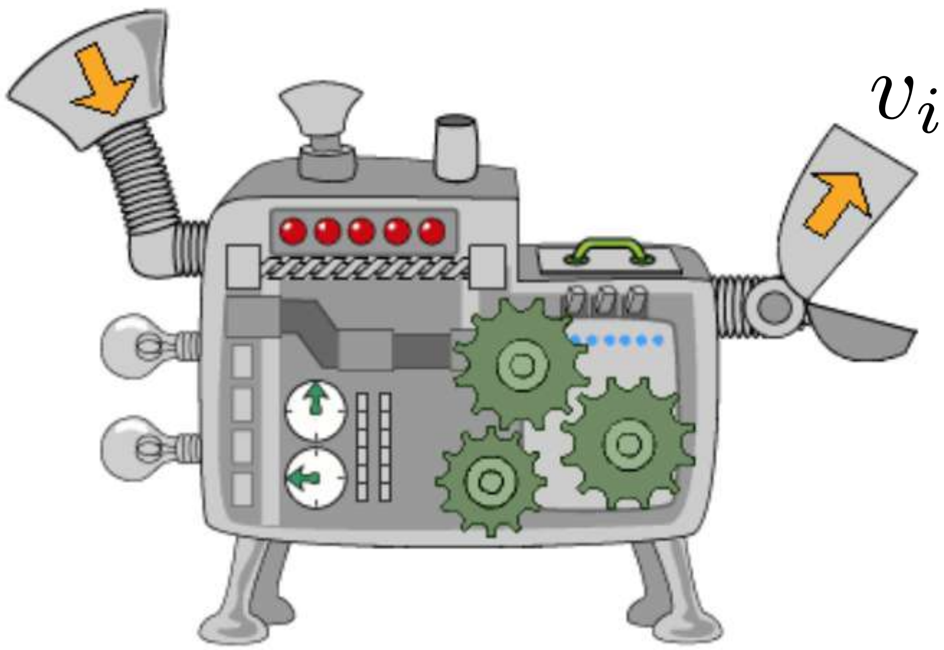
Rosenblatt (1957)

Mathematica 10+

Schematic representation of feedforward neural network. The top figure denotes the perceptron (a single neuron), the bottom, the multiple neurons and multiple layers of the neural network.

Neural Network

J_i



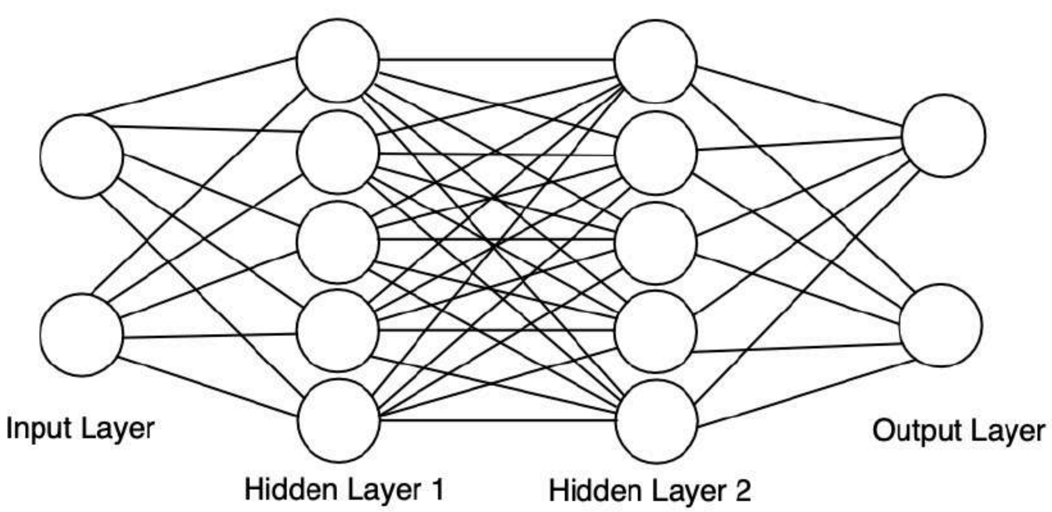
v_i

$$\{J_1, \ldots, J_n\} \longrightarrow \{v_1, \ldots, v_n\}$$

$$J_i \in T$$

$$\{J'_1, \ldots, J'_m\} \longrightarrow ???$$

$$J'_i \in T^c$$



$\vec{J}$ $\quad 100 \times 18 \quad 100 \times 100 \quad \sum_{a=1}^{100}$

12000 hyperparameters

Jones polynomials are represented as 18-vectors

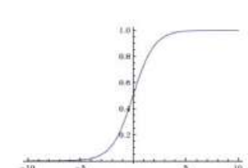
$$\vec{J}_K = (\text{min}, \text{max}, \text{coeffs}, 0, \ldots, 0)$$

Two layer neural network in **Mathematica**

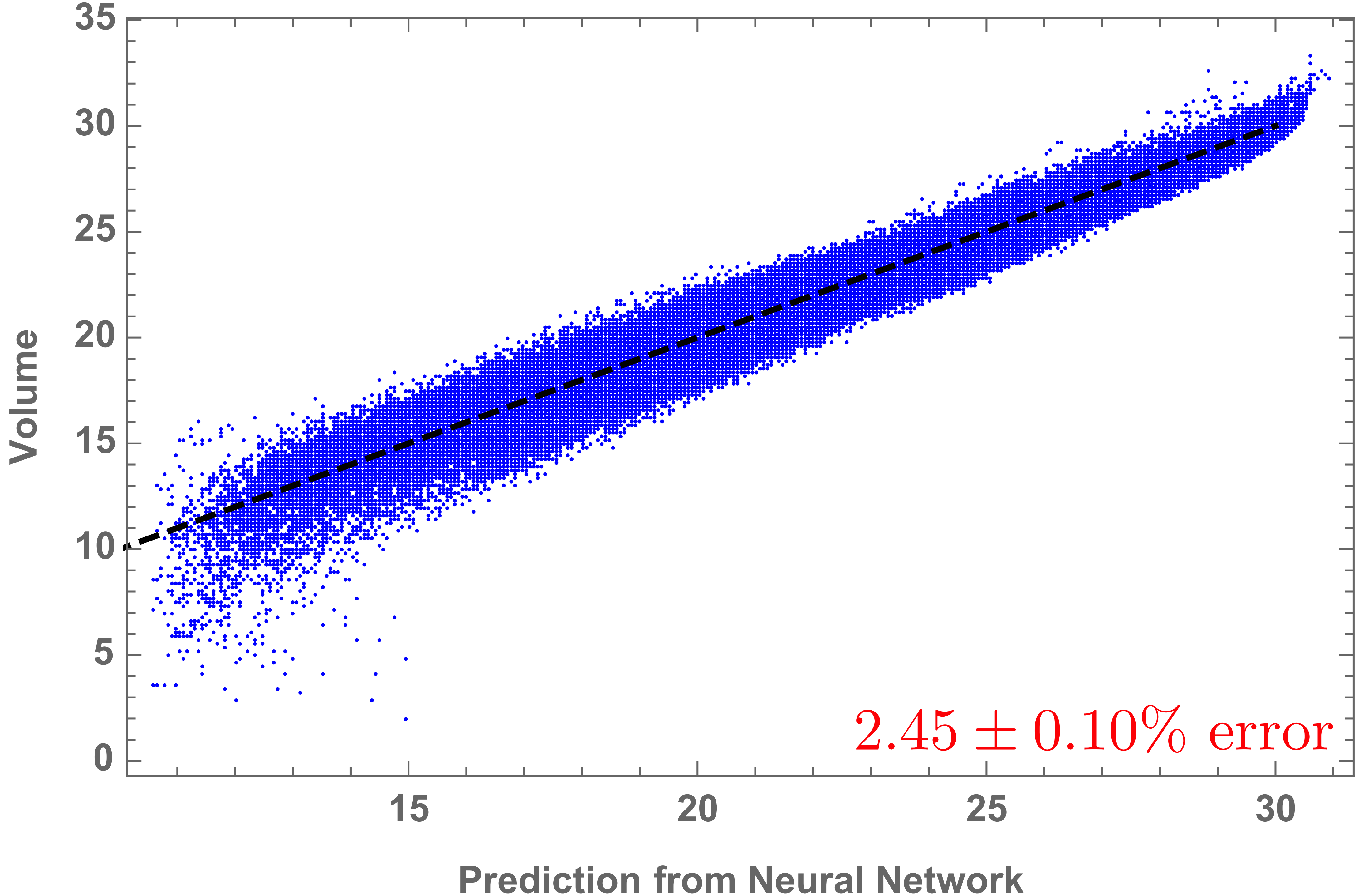
$$f_\theta(\vec{J}_K) = \sum_a \sigma\left(W_\theta^2 \cdot \sigma(W_\theta^1 \cdot \vec{J}_K + \vec{b}_\theta^1) + \vec{b}_\theta^2\right)^a$$

Logistic sigmoids for the hidden layers

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

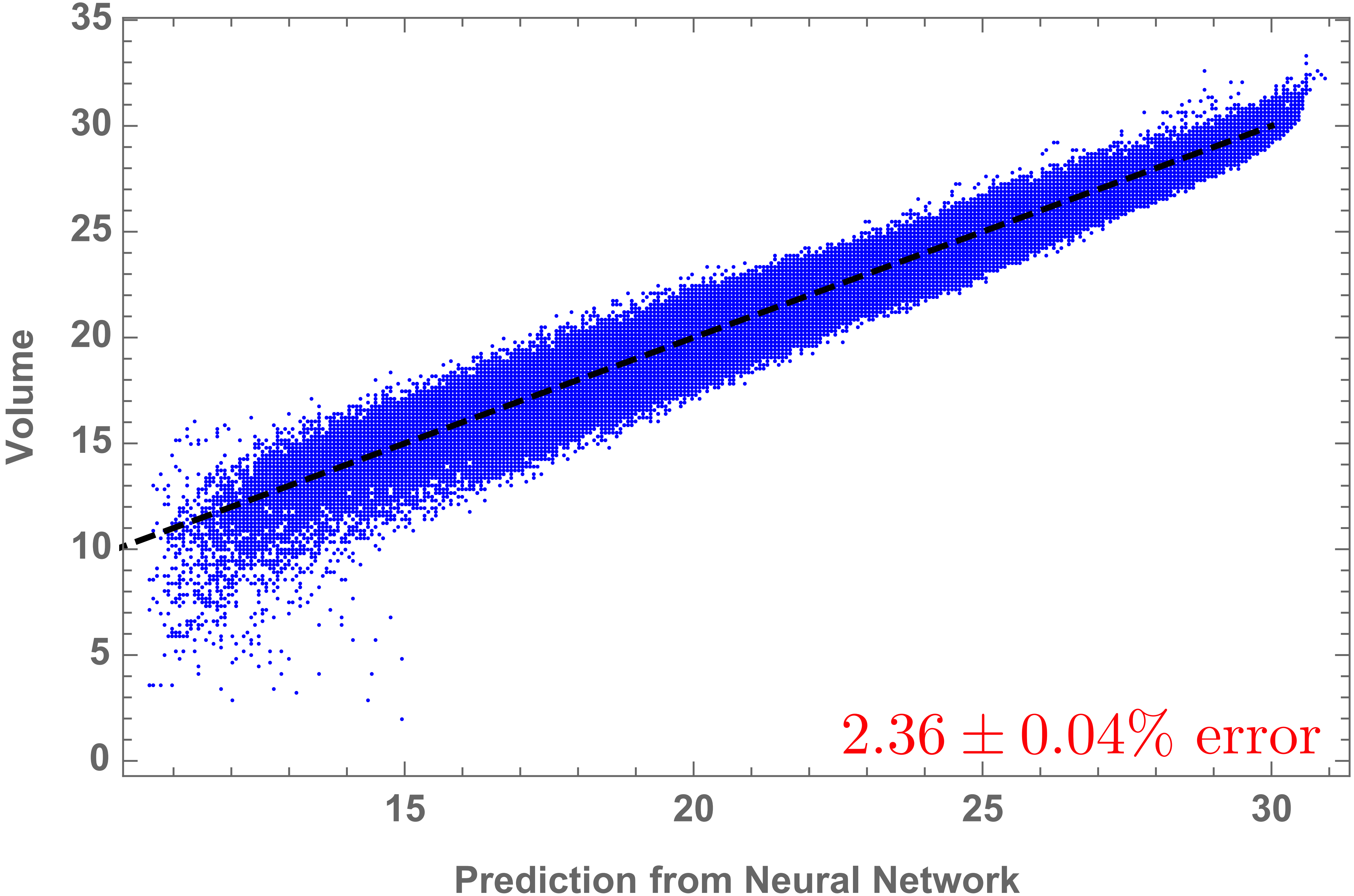


Neural Network



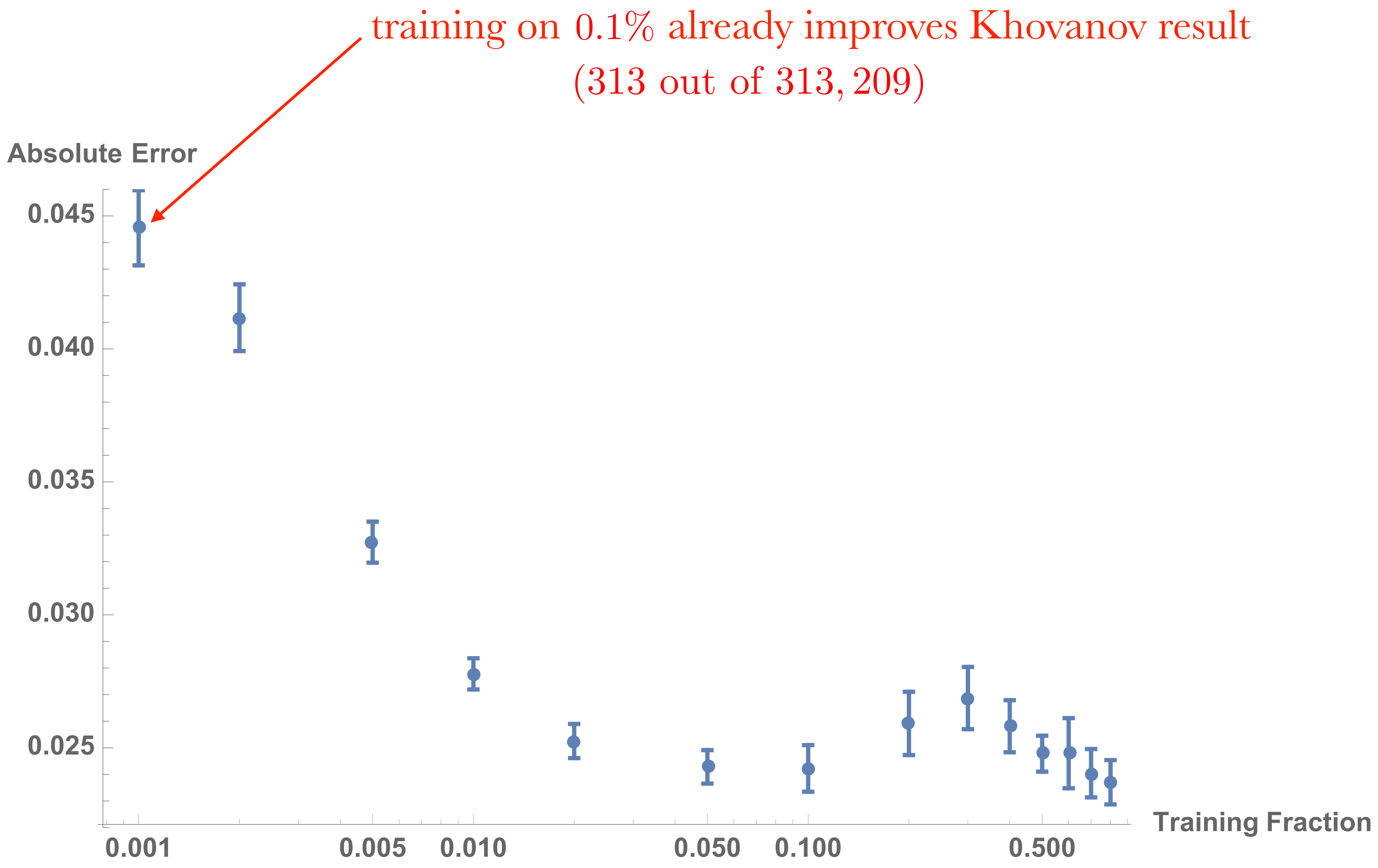
trained on 10% of the 313, 209 knots up to 15 crossings

Neural Network



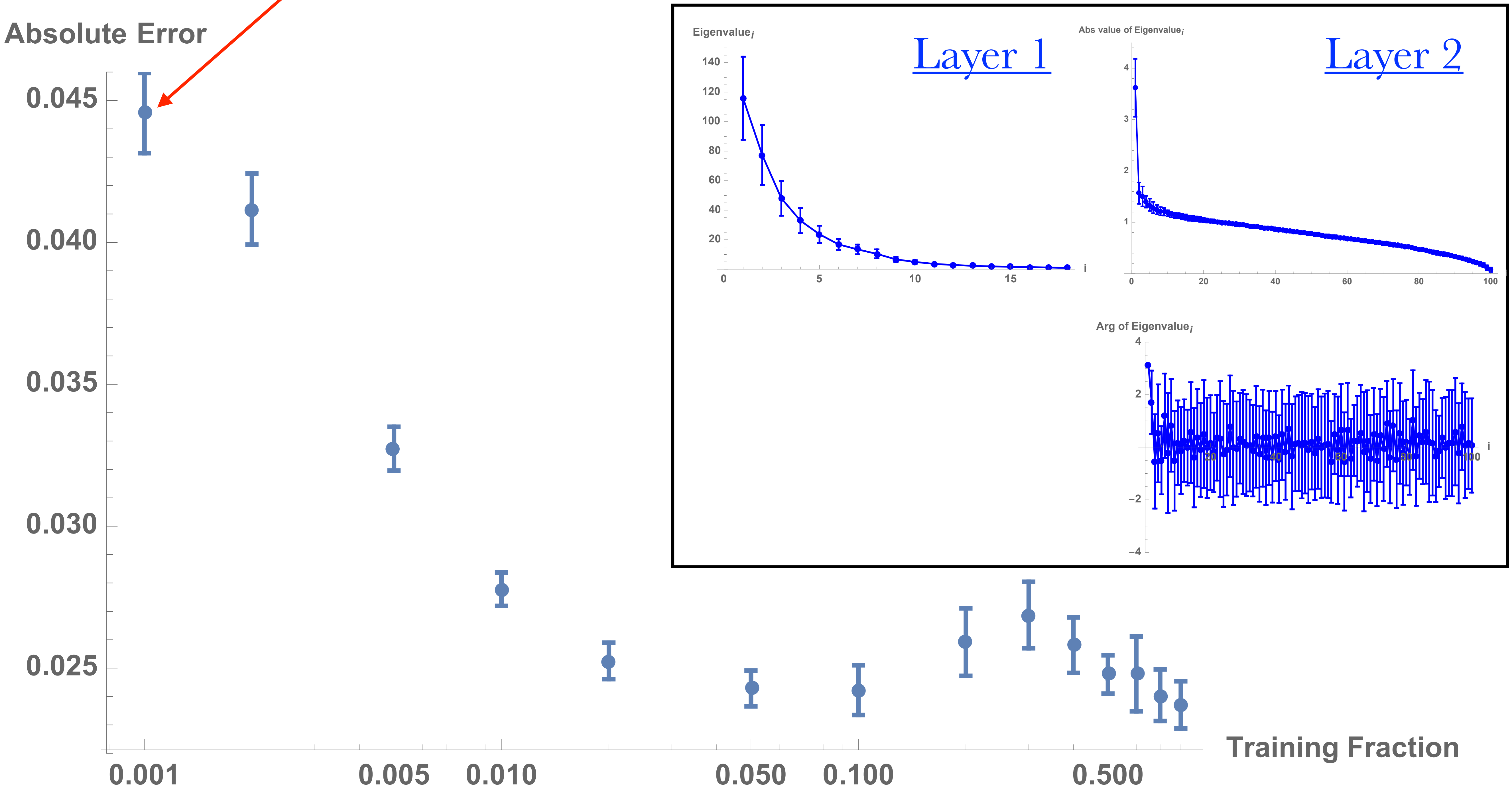
trained on 10% of the 313, 209 knots up to 15 crossings

Learning Curve



Learning Curve

training on 0.1% already improves Khovanov result
(313 out of 313, 209)



Result

$$v_i = f(J_i) + \text{small corrections}$$

Result

$$v_i = f(J_i) + \text{small corrections}$$

- J_i does not uniquely identify a knot

  e.g., 4_1 and K11n19 have same Jones polynomial, different volumes

- $174,619$ unique Jones polynomials

  2.83% average spread in volumes for a Jones polynomial
  intrinsic mitigation against overfitting

- See same behavior with 1,701,903 knots up to 16 crossings

  (database compiled from **Knot Atlas** and **SnapPy**)

Result

$$v_i = f(J_i) + \text{small corrections}$$

- Neural network does better than more refined topological invariants

- Beyond the volume conjecture in Chern–Simons $\left\{ \begin{array}{l} \text{weak coupling limit of } SL(2,\mathbb{C}) \text{ Chern–Simons} \\ \text{strong coupling limit of } SU(2) \end{array} \right.$

  Jones polynomial (quantum) $\longleftrightarrow$ volume (classical)

- Failed experiments (*e.g.*, learning Chern–Simons invariant) also teach us something — maybe about the need for underlying homology theory

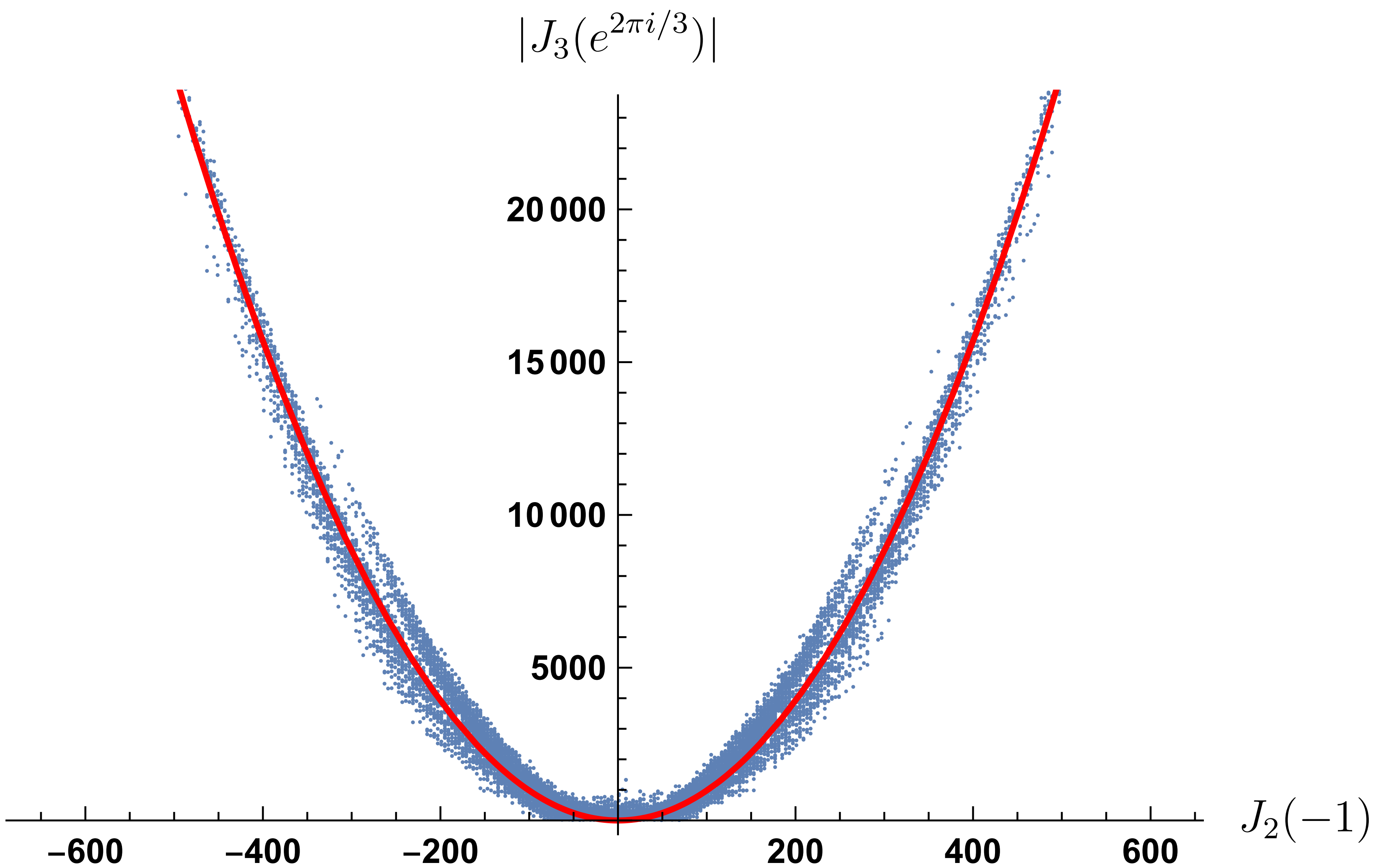
$$\lim_{N\to\infty} \frac{2\pi \log J_{K,N}(e^{\frac{2\pi i}{N}})}{N} = \mathrm{Vol}(K) + 2\pi^2 i \mathrm{CS}(K)$$

cf. Calabi–Yau Hodge numbers, line bundle cohomology, etc.

- We want a **not** machine learning knot result

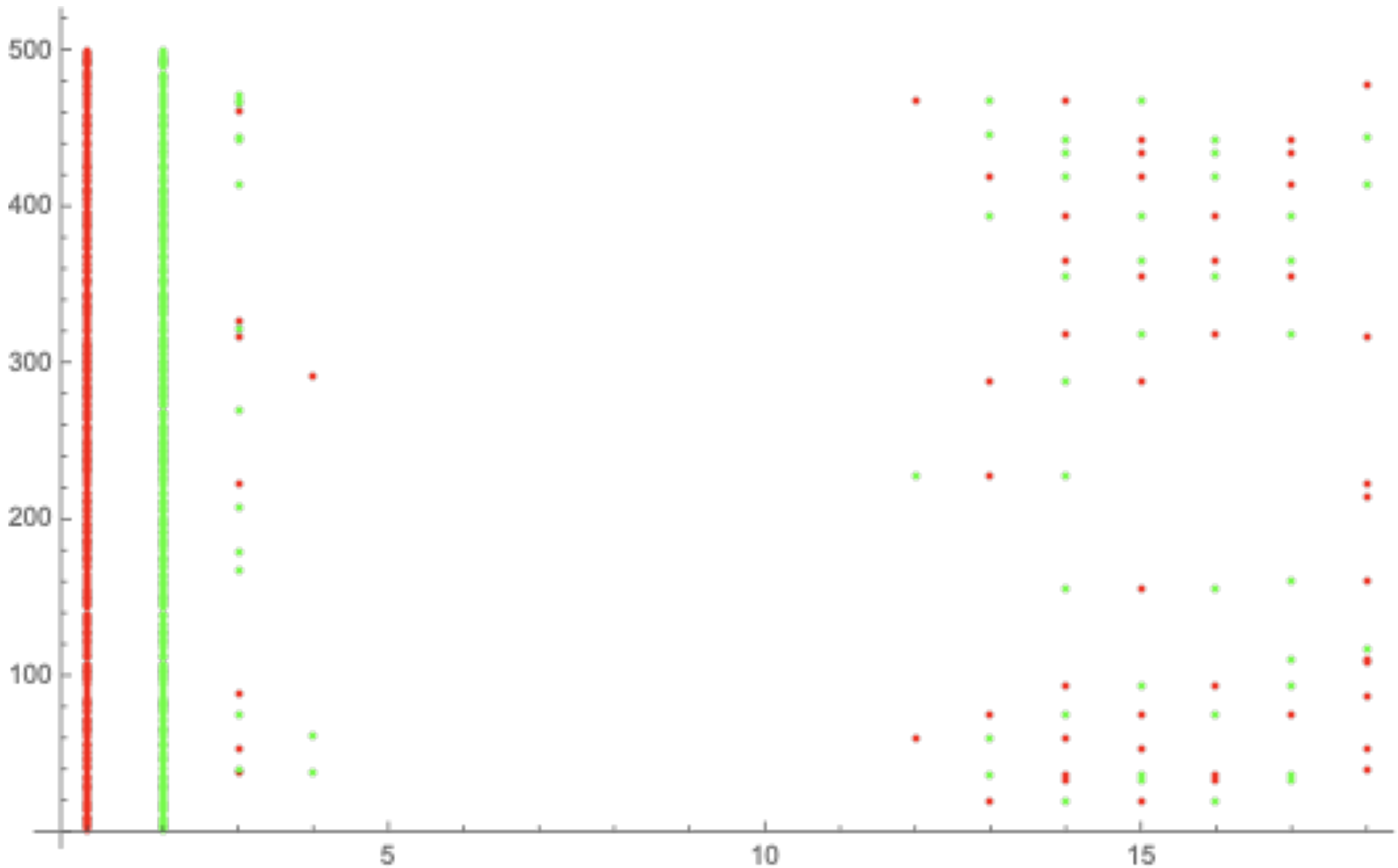
Towards the Volume Conjecture

- The Volume Conjecture: $\lim_{N\to\infty} \frac{2\pi \log |J_{K,N}(e^{\frac{2\pi i}{N}})|}{N} = \mathrm{Vol}(K)$



- 11,921 colored Jones polynomials at $N = 3$

Principal Component Analysis



- One layer network — error is only slightly worse: $2.45 \pm 0.04\%$ error
- Where is the weight matrix most dense? — 500 neurons

No Degrees Needed

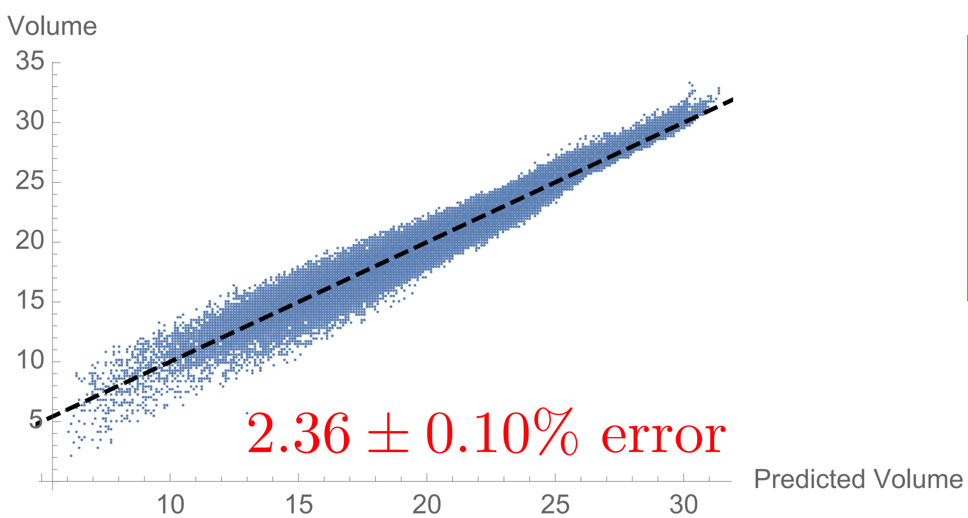
- Jones polynomials were fed in to the neural network as 18-vectors

$$\vec{J}_K = (\text{min}, \text{max}, \text{coeffs}, 0, \ldots, 0)$$

- Suppose we drop the degrees so that the Jones polynomial is no longer recoverable from the input vector

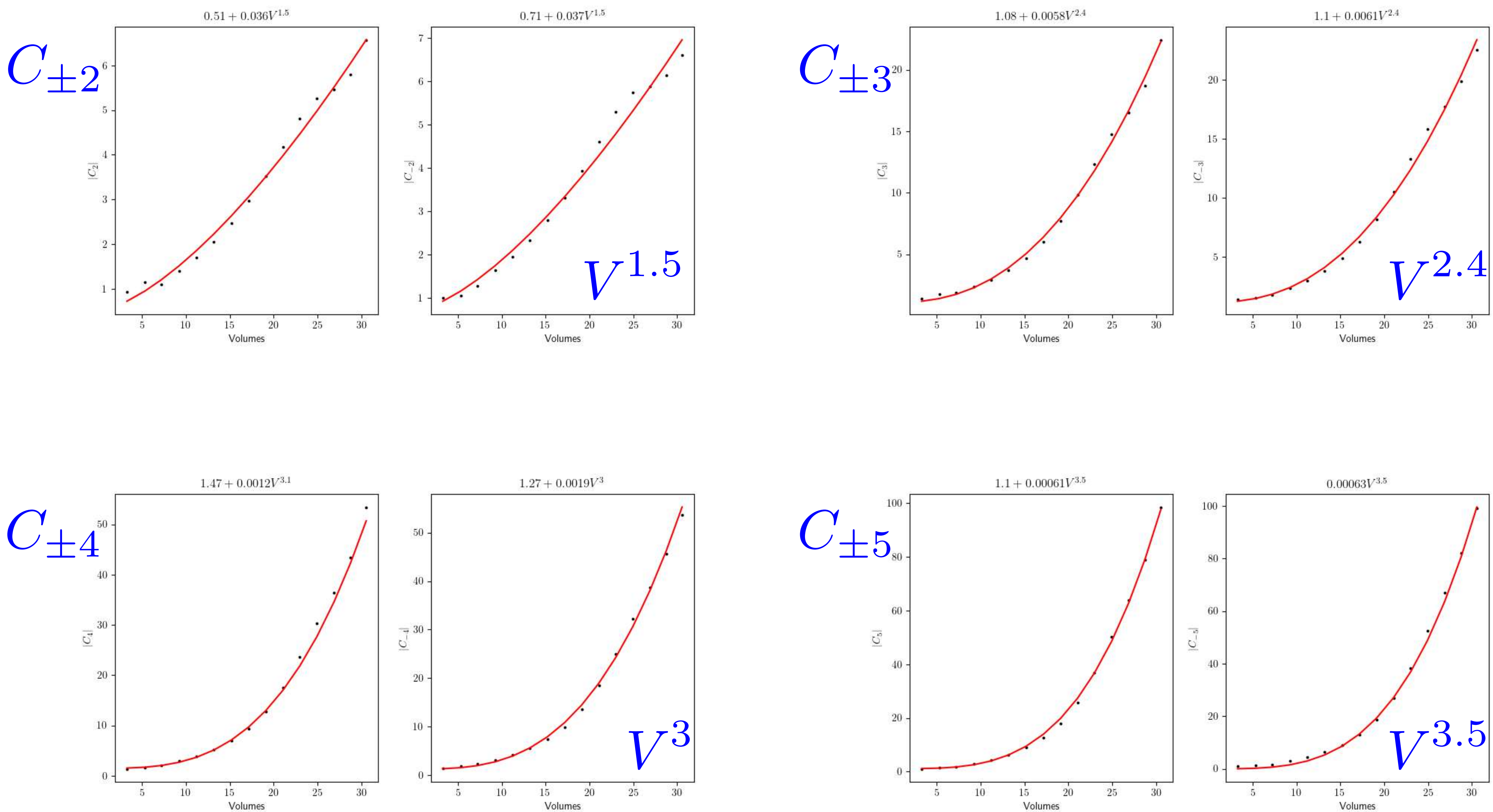
$$\vec{C}_K = (\text{coeffs}, 0, \ldots, 0)$$

- Results are unchanged!



Volume is determined only by the coefficients in the Jones polynomial

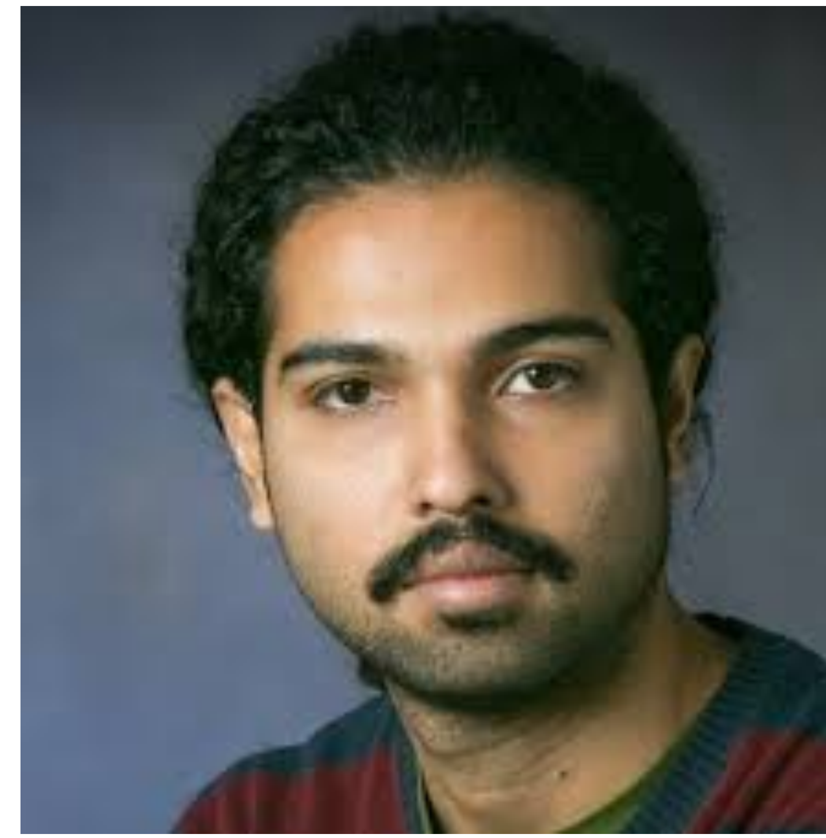
Coefficients



Collaborators

Jessica Craven

Arjun Kar



Onkar Parrikar

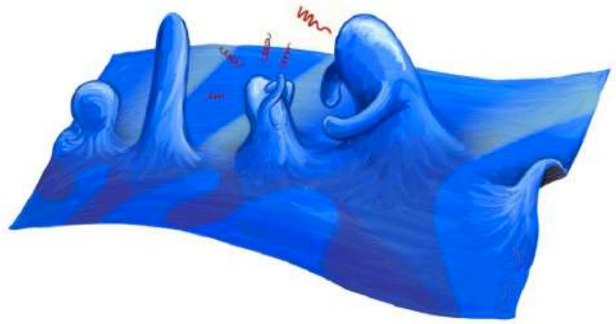
arXiv:1902.05547, 2010.nnnnn

Masses in QCD

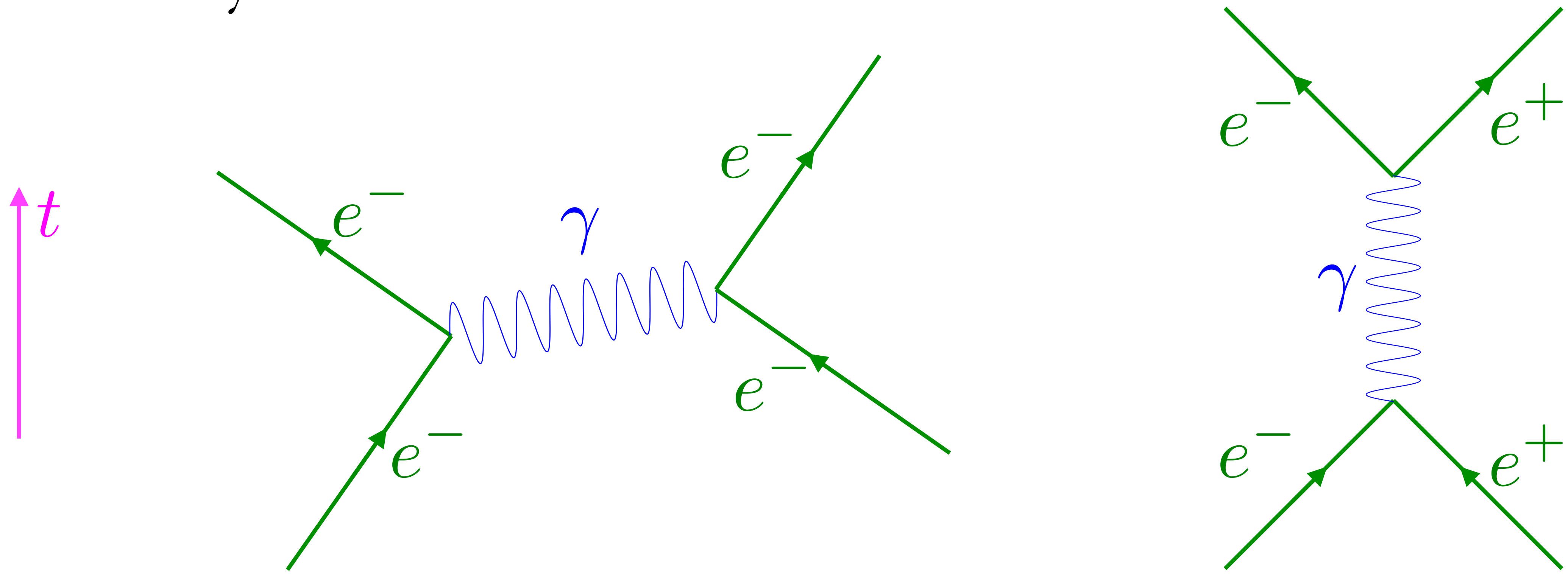
Quantum Field Theory

- A field has a value at every point in spacetime

- Particles are local excitations of these fields



- To define a quantum field theory, we must specify the fields and how they interact



- Electrons and positrons interact by exchanging photons, for example

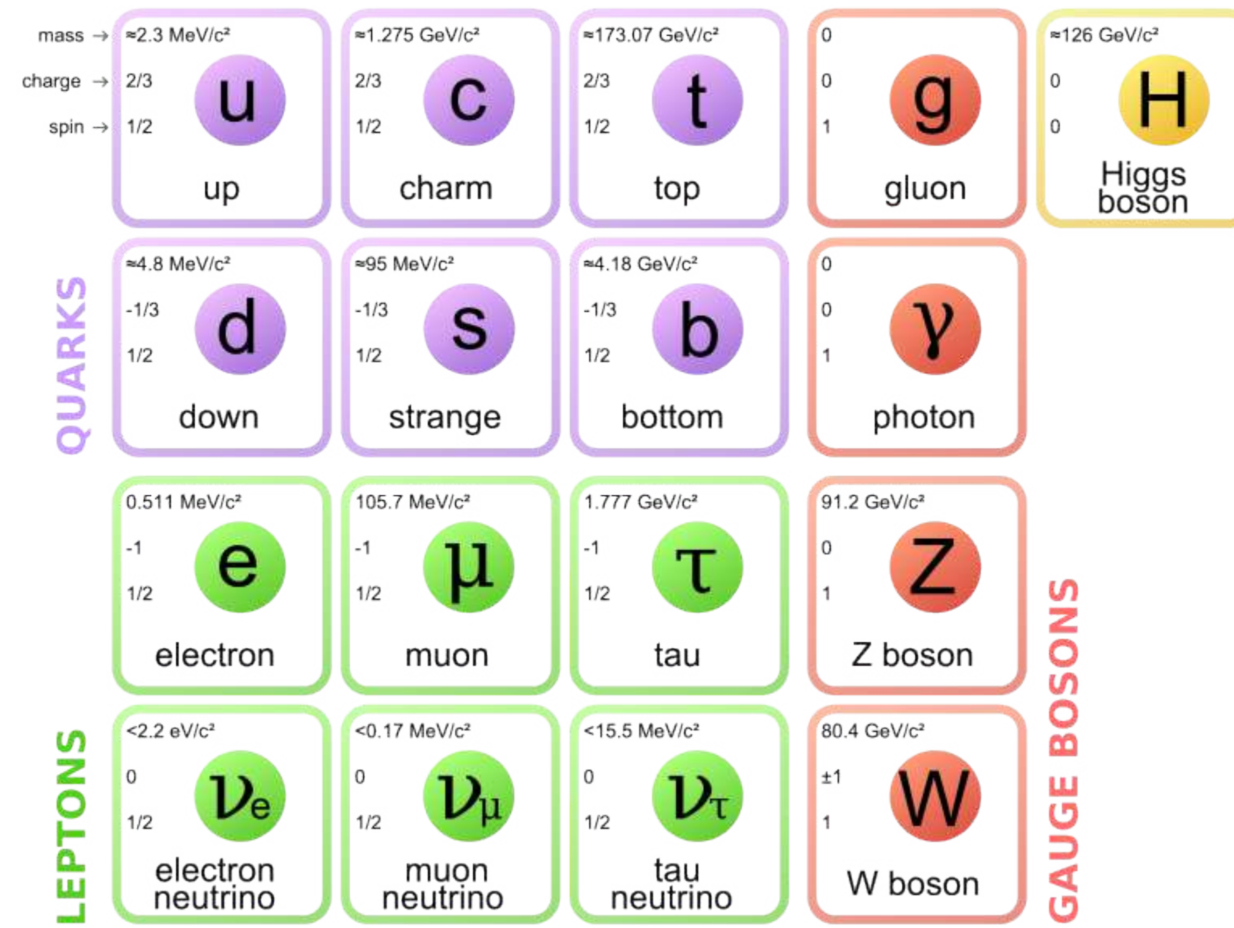
Quantum Field Theory

- Fundamental forces are described by quantum field theory

- Standard Model

  electromagnetism
  weak force
  strong force
  Higgs effect

  everything except gravity



- $\alpha_{\text{exp}}^{-1} = 137.035999139(31)$

  $\alpha_{\text{th}}^{-1} = 137.035999173(35)$

LHC has not seen BSM physics
Nevertheless, SM is incomplete

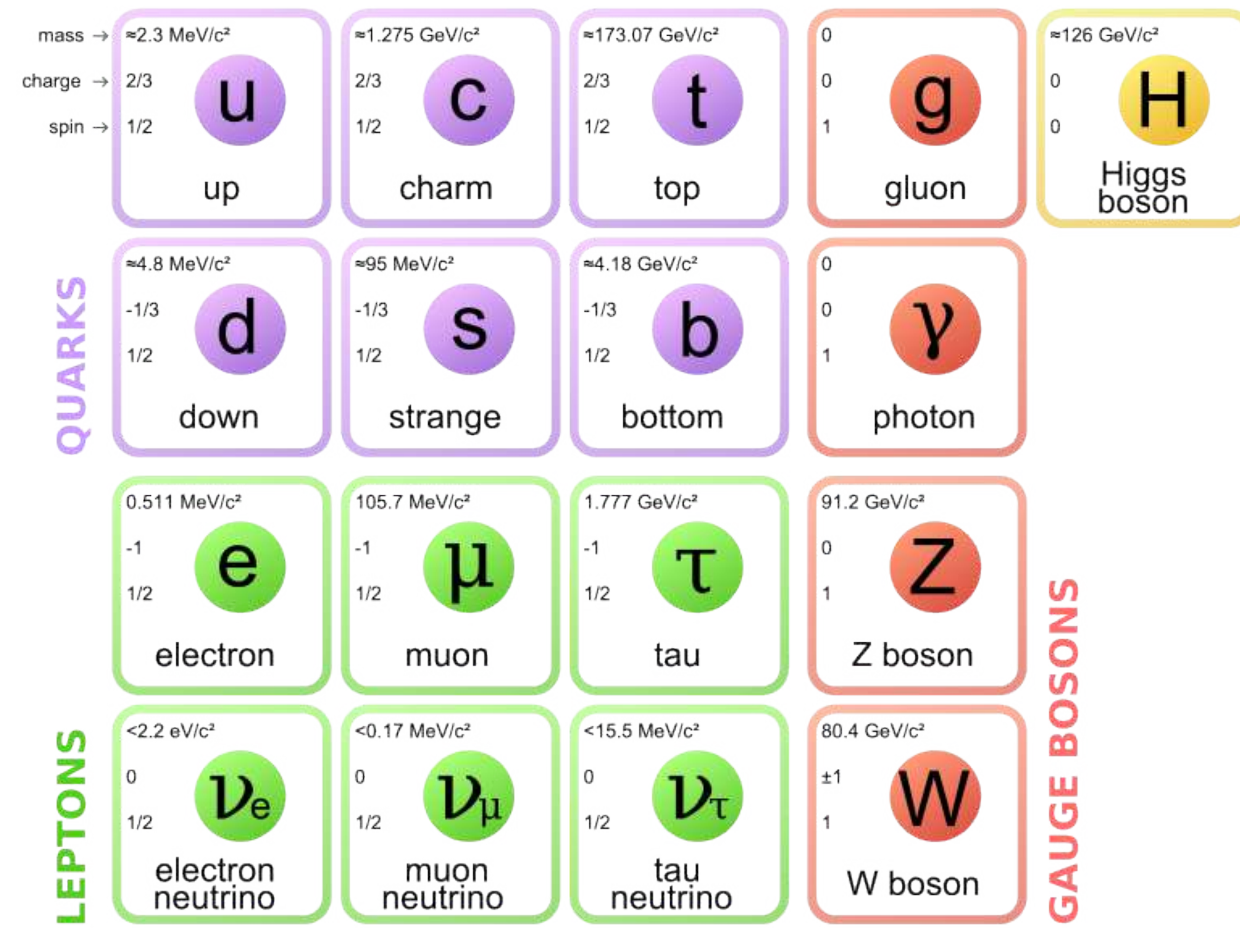
Quantum Field Theory

- Fundamental forces are described by quantum field theory

- Standard Model

electromagnetism
weak force
strong force ⟵ QCD
Higgs effect

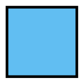
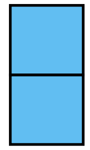
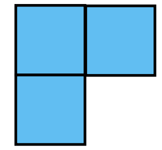
everything except gravity



- $\alpha_{\rm exp}^{-1} = 137.035999139(31)$
  $\alpha_{\rm th}^{-1} = 137.035999173(35)$

LHC has not seen BSM physics
Nevertheless, SM is incomplete

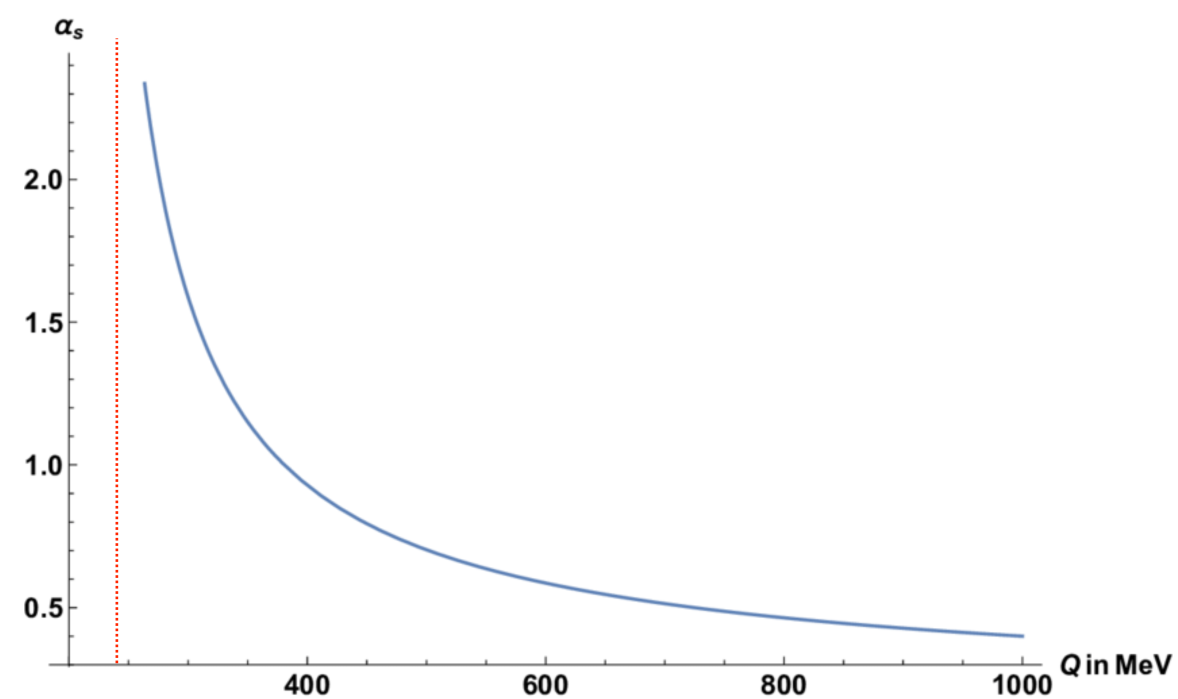
Confinement

- $q \in$  , $\bar{q} \in$  , $g \in$  of $SU(3)$ QCD

- Asymptotic freedom

$$\beta(g_s) = \frac{\partial g_s}{\partial \log \mu} = -\frac{g_s^3}{16\pi^2}\left(\frac{11}{3}N - \frac{2}{3}N_f - \frac{1}{3}N_s\right)$$

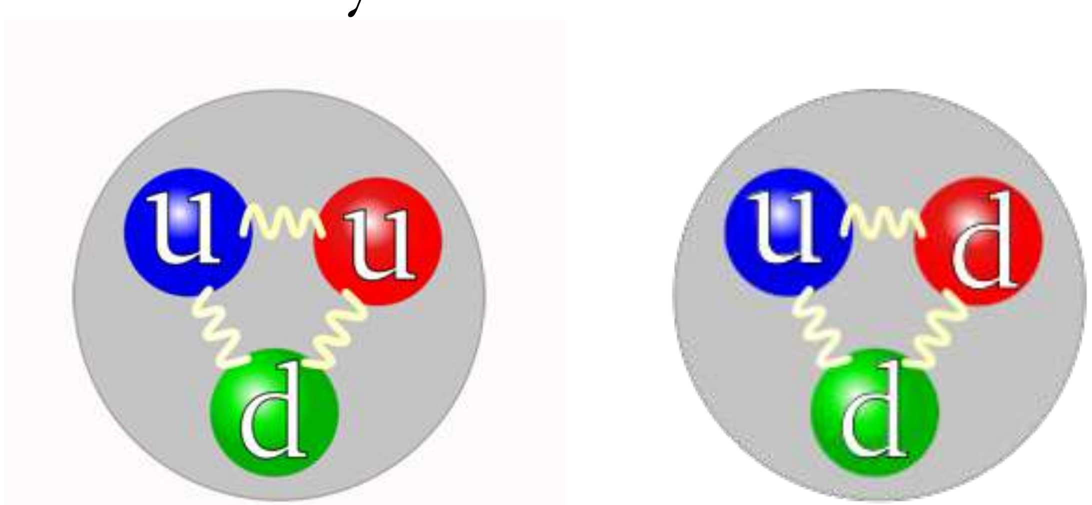
$$\alpha_s(k^2) = \frac{g_s^2(k^2)}{4\pi}$$



$$\Lambda_{\rm QCD} = 218 \pm 24 \text{ MeV}$$

Gross, Wilczek (1973)
Politzer (1973)

- Infrared slavery

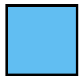
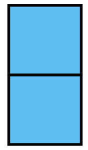
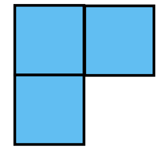


proton neutron

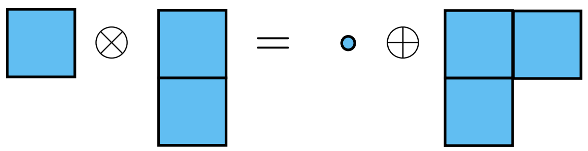
$$d, s, b: \ Q = -\frac{1}{3}$$

$$u, c, t: \ Q = +\frac{2}{3}$$

Hadronic Spectrum

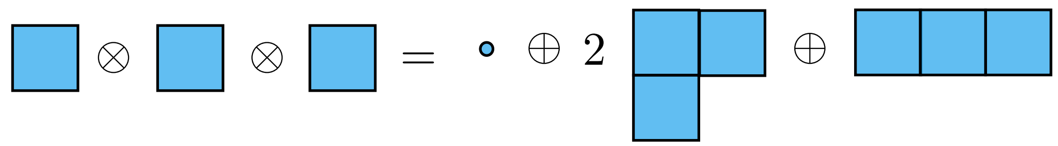
- $q \in$  , $\bar{q} \in$  , $g \in$  of $SU(3)$ QCD

- Mesons and baryons



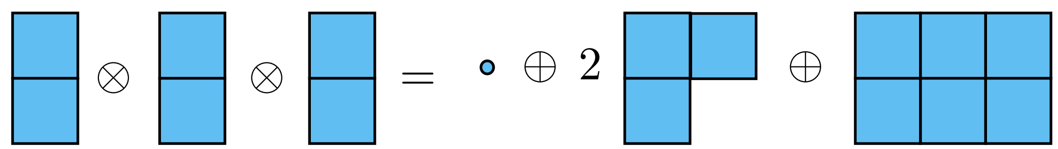
$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$$

$$q_i^a \bar{q}_{a,j}$$



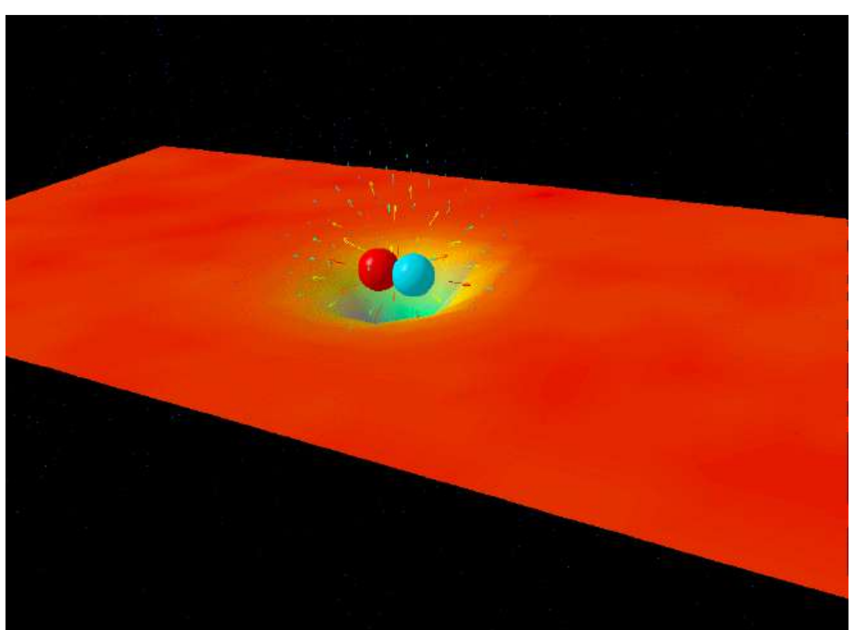
$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus 2 \times \mathbf{8} \oplus \mathbf{10}$$

$$\epsilon_{abc} q_i^a q_j^b q_k^c$$



$$\overline{\mathbf{3}} \otimes \overline{\mathbf{3}} \otimes \overline{\mathbf{3}} \quad \mathbf{1} \oplus 2 \times \mathbf{8} \oplus \overline{\mathbf{10}}$$

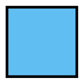
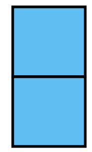
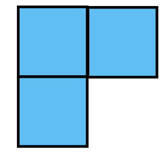
$$\epsilon^{abc} \bar{q}_{a,i} \bar{q}_{b,j} \bar{q}_{c,k}$$



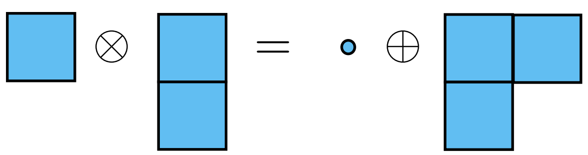
meson in QCD
quark/antiquark pair
held together by gluons

} **color confinement**

Hadronic Spectrum

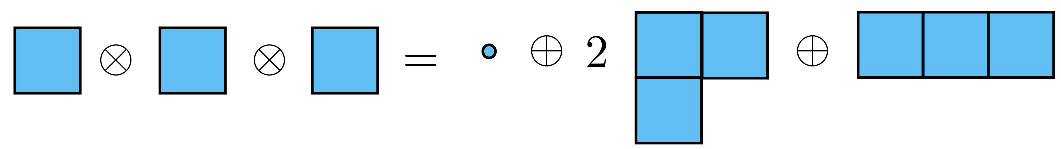
- $q \in$  , $\bar{q} \in$  , $g \in$  of $SU(3)$ QCD

- Mesons and baryons



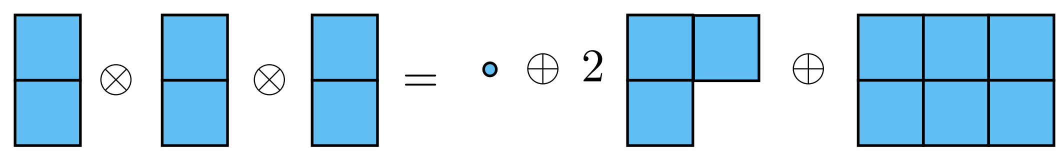
$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$

$q_i^a \bar{q}_{a,j}$



$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus 2 \times \mathbf{8} \oplus \mathbf{10}$

$\epsilon_{abc} q_i^a q_j^b q_k^c$



$\bar{\mathbf{3}} \otimes \bar{\mathbf{3}} \otimes \bar{\mathbf{3}} \quad \mathbf{1} \oplus 2 \times \mathbf{8} \oplus \overline{\mathbf{10}}$

$\epsilon^{abc} \bar{q}_{a,i} \bar{q}_{b,j} \bar{q}_{c,k}$

- Most of hadron's mass from strong coupling effects (*viz.*, confinement)

- In large-N, baryons are solitons in meson spectrum

  mesons essentially free particles with couplings $\mathcal{O}(\frac{1}{N})$

  baryons monopole-like excitations with masses $\mathcal{O}(N)$

Witten (1979)

Hadronic Spectrum

- 196 mesons, 43 baryons — mesons are bosons, baryons are fermions
  (integer spin) (half-integer spin)

- Valence quark content from Monte Carlo analysis described in Particle Data Book

- Include: I , J , P ; do not include G , C , width due to elisions

- Optimize architecture of neural network based on training with 80% of mesons to predict masses

- Test the neural network on baryons

- Predict tetraquark, pentaquark masses; discriminate composition hypotheses

AdS/ML

- Beautiful recent work by Hashimoto, *et al.*

- Train on lattice data for chiral condensate vev vs. quark mass

- Obtain metric from parameters in neural network and follow AdS/QCD dictionary despite not being in large-N limit

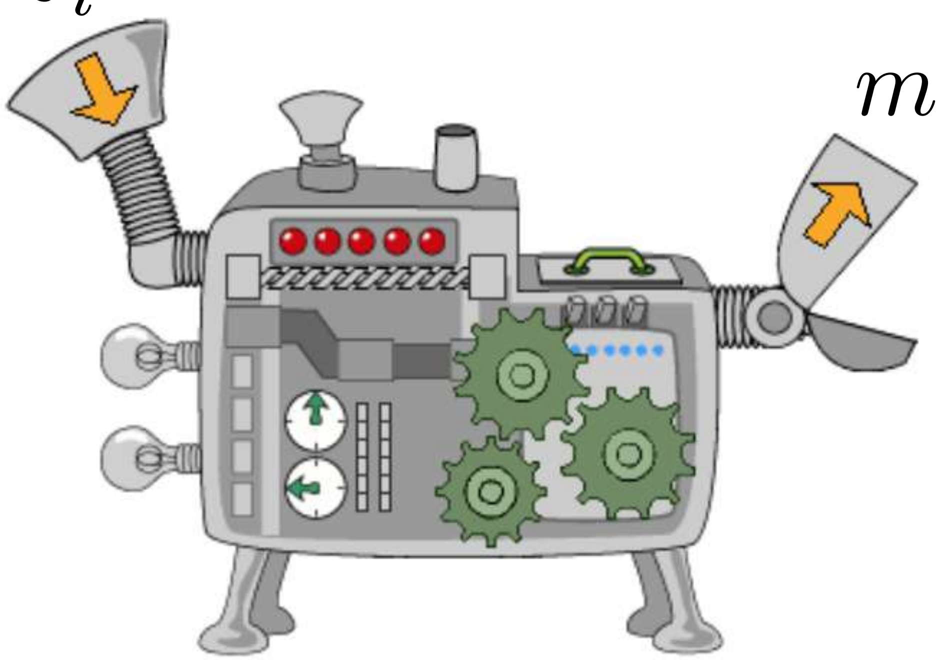
- Predict form of quark–antiquark potential

Hashimoto, Sugishita, Tanaka, Tomiya (2018)
Hashimoto (2019)
Akutagawa, Hashimoto, Sugimoto (2020)

- We do something similar: we train on one data set and predict on a different data set

Neural Network

$\vec{v}_i$ m_i

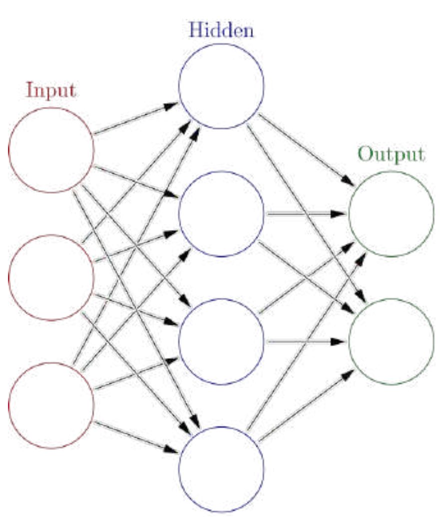


$$\{\vec{v}_1, \ldots, \vec{v}_n\} \longrightarrow \{m_1, \ldots m_n\}$$

fix architecture training on 80% of mesons

$$\{\vec{v}_1', \ldots, \vec{v}_\ell'\} \longrightarrow ???$$

baryons for testing, predict pentaquarks, etc.



$\vec{v}$ $\quad 50 \times 13 \quad \sum_{a=1}^{50}$

700 hyperparameters

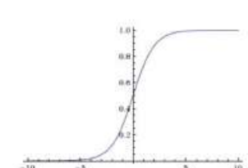
hadrons as 13-vectors

$$\vec{v} = (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P)$$

Monolayer neural network in Mathematica

Logistic sigmoids for the hidden layer

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



Gaussian Process

- Stochastic process such that every finite collection of random variables has a multivariate normal distribution

- Can be used as a prior probability distribution in Bayesian inference

- Interpolation and statistical inference via **kriging** or **Wiener—Kolmogorov prediction**

- Infinite width limit of single hidden layer, fully connected neural network

Neal (1994)

Gaussian Process

- Choose **kernel** $K_{ij} := k(x_i, x_j)$

$$k_{\mathrm{SE}}(x_i, x_j) \;=\; \sigma_f^2 \exp\left(-\frac{1}{2}(x_i - x_j)^{\mathrm{T}} \Lambda^{-1} \, (x_i - x_j) \right)$$

$$k_{\mathrm{RQ}}(x_i, x_j) \;=\; \sigma_f^2 \left(1 + \frac{1}{2\alpha}(x_i - x_j)^{\mathrm{T}} \Lambda^{-1} \, (x_i - x_j)\right)^{-\alpha}$$

diagonal matrix

- Define **covariance** $\mathrm{cov}(x_i, x_j) := k(x_i, x_j) + N_{ij}$, $N_{ij} = \sigma_n^2 \delta_{ij}$

- Let $x_i \mapsto y_i$, $i = 1, \ldots, M$

- Maximize **log marginal likelihood** to fix hyperparameters

$$\log p(y|X) := -\frac{1}{2} y^T (K+N)^{-1} y - \frac{1}{2} \log |K+N| - \frac{M}{2} \log(2\pi)$$

small data

Gaussian Process

- Let $x_\star$ be a new vector; $k_{\star,i} = k(x_\star, x_i)$ defines an M-vector

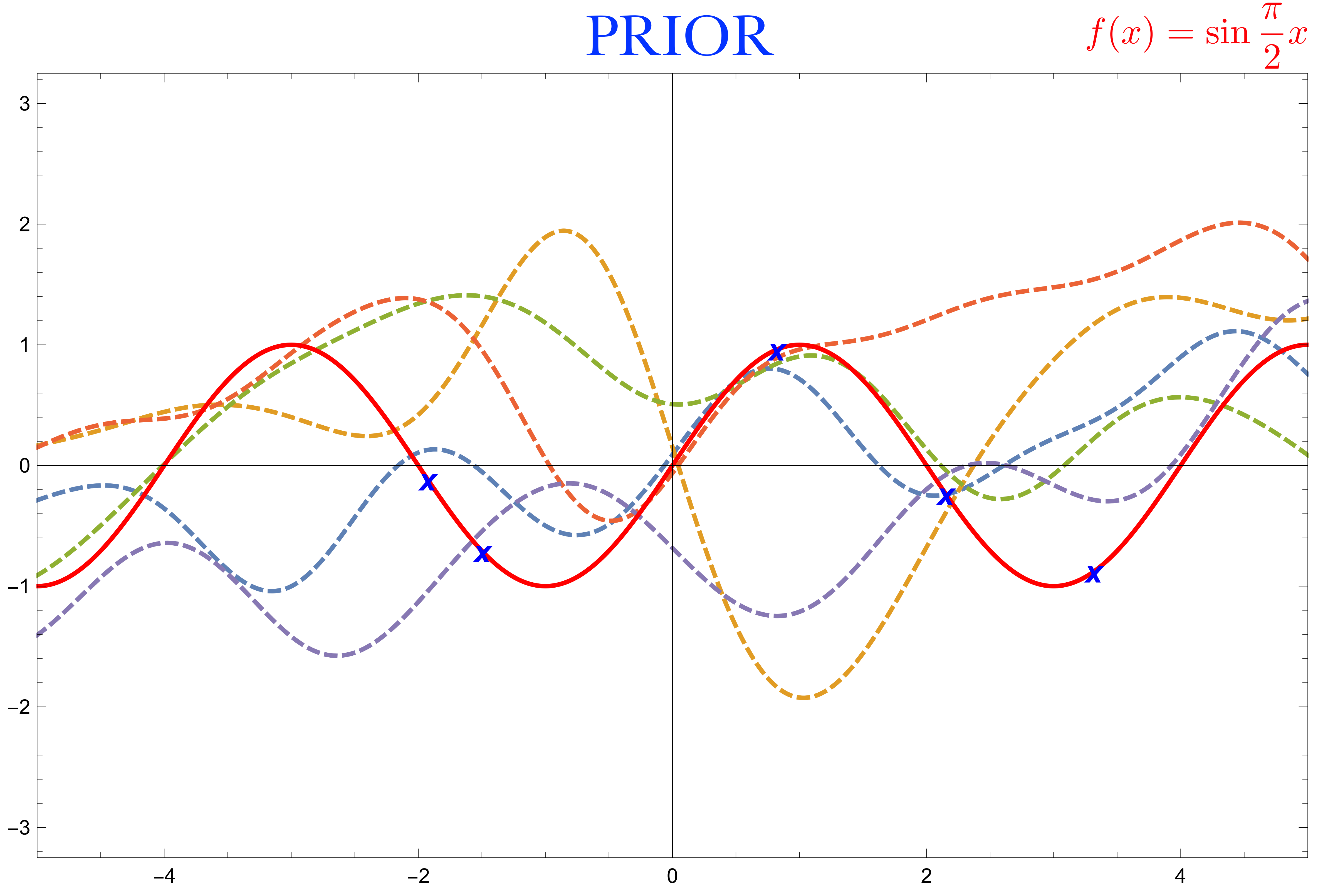
- We predict the **mean** and **variance**

$$\mu_\star = {k_\star}^T (K+N)^{-1} y$$

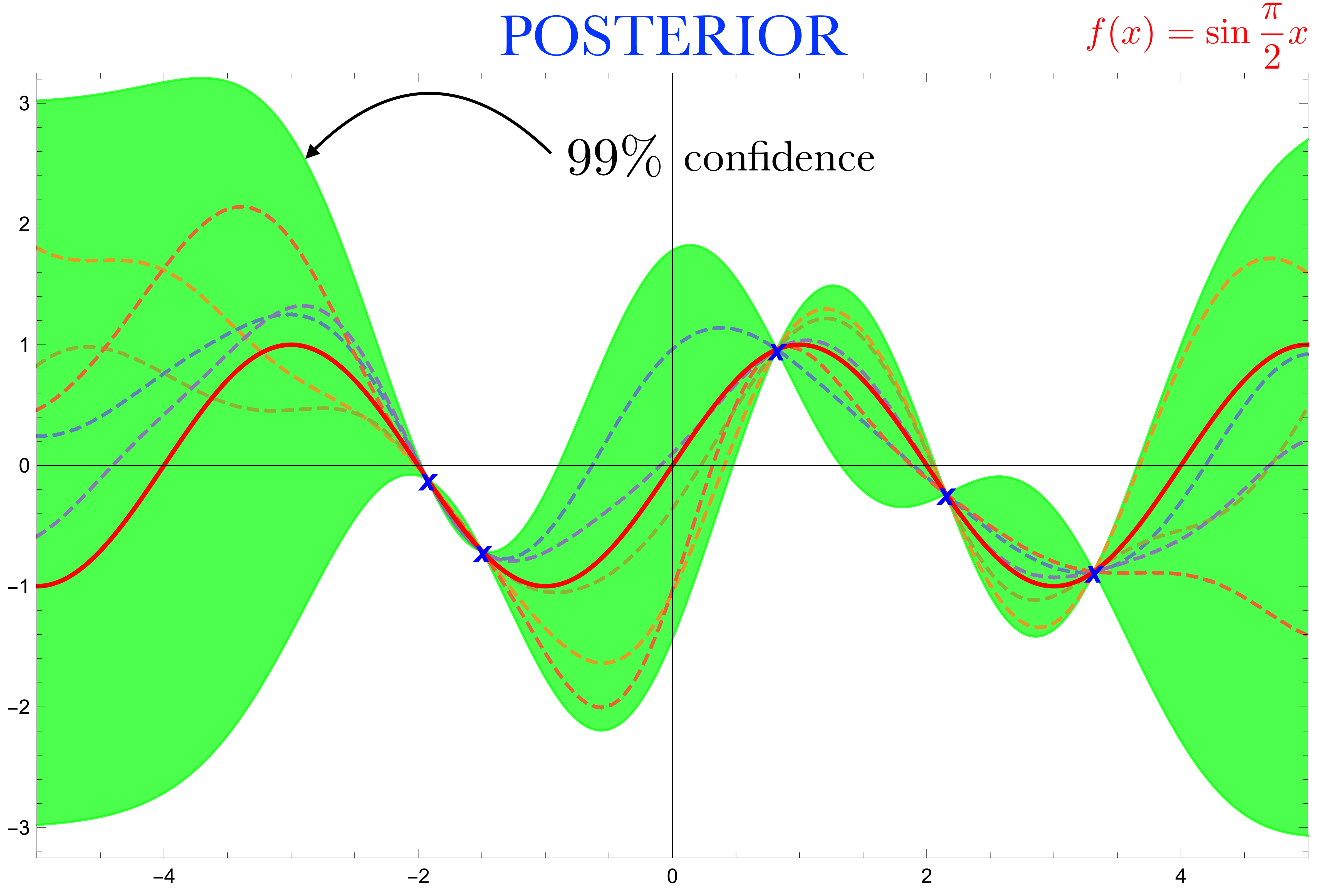
$$\text{var}_\star = k(x_\star, x_\star) - k_\star^T (K+N)^{-1} k_\star$$

- Implemented with GPML 4.2 in Matlab R2019_b

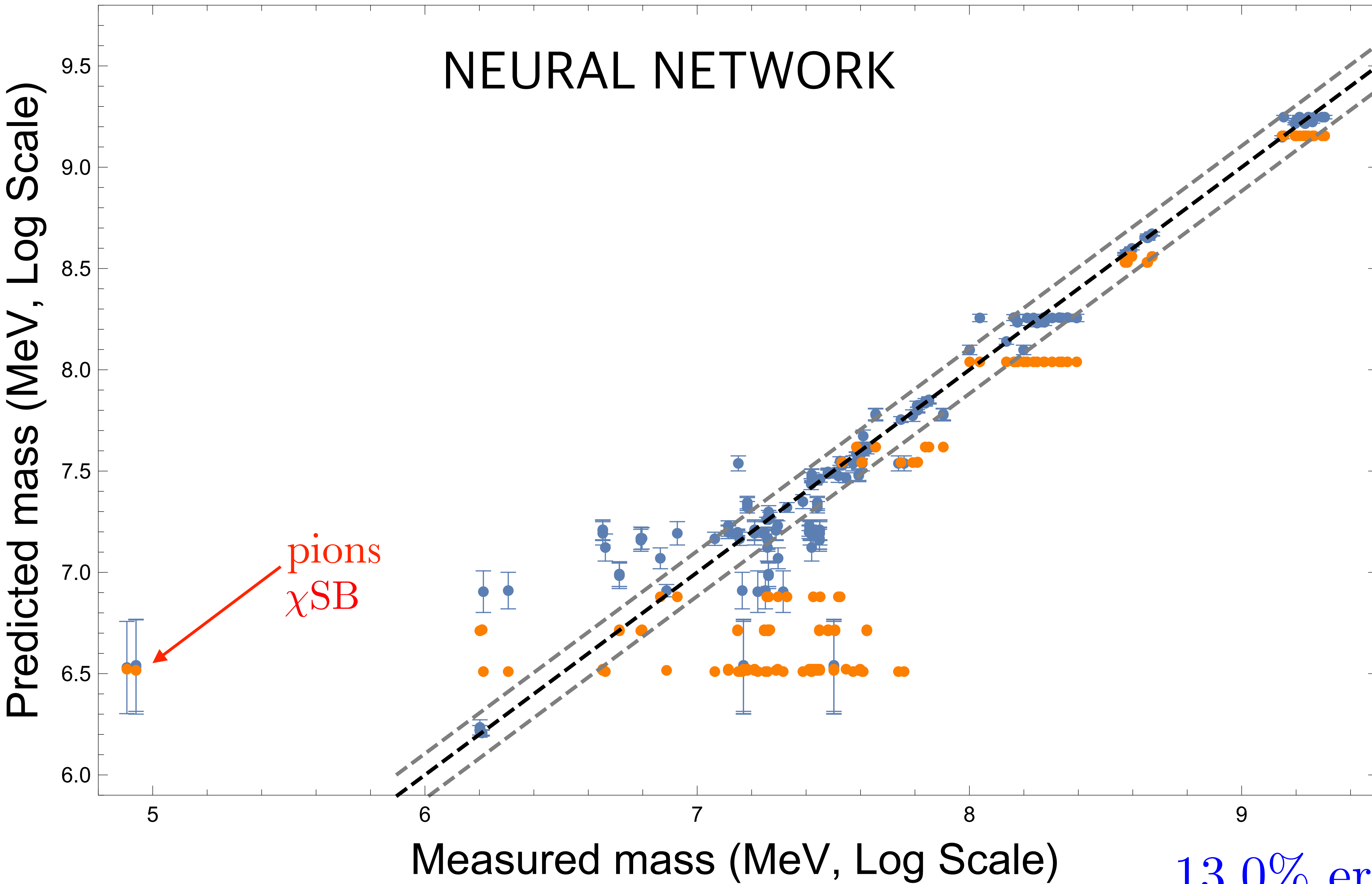
Gaussian Process



Gaussian Process

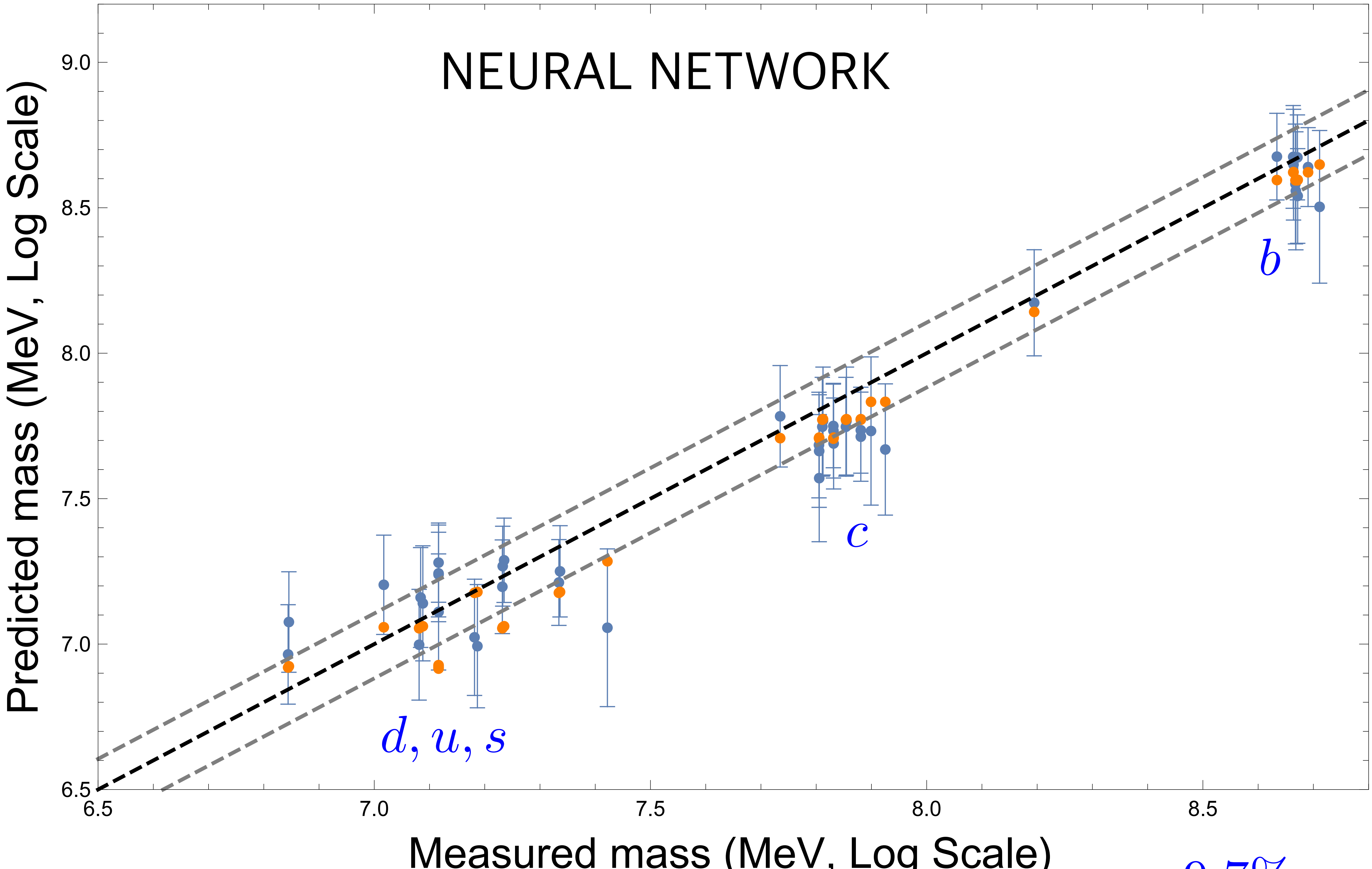


Mesons



13.0% error

Baryons

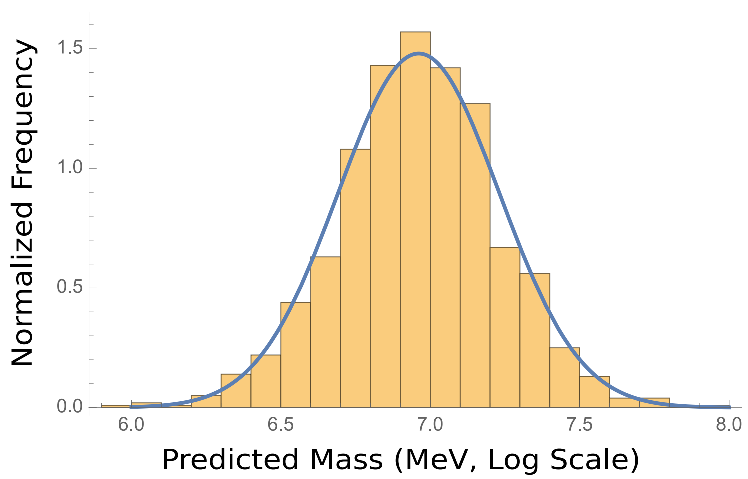


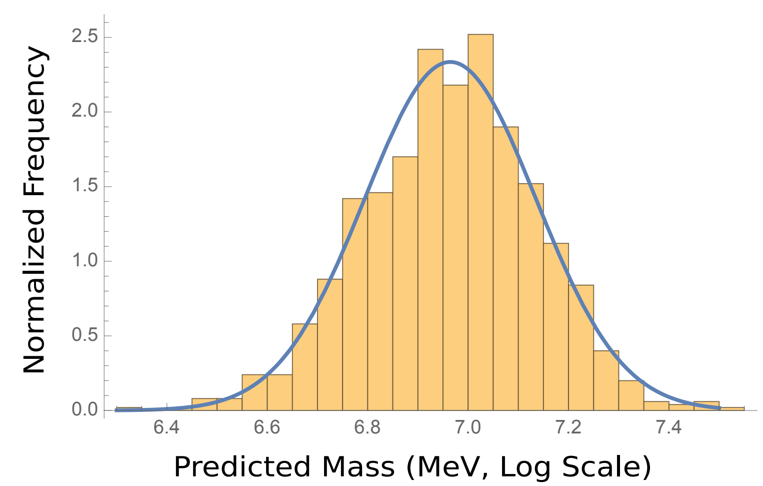
9.7% error

Error Analysis

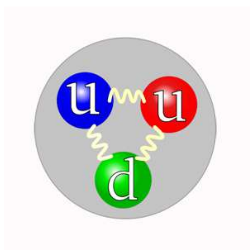
- We do $\vec{v}_i \mapsto \log m_i$; effectively divide mass by 1 MeV
  training vectors do not uniquely identify a meson

- Posterior distribution has mean μ and variance σ





m_p

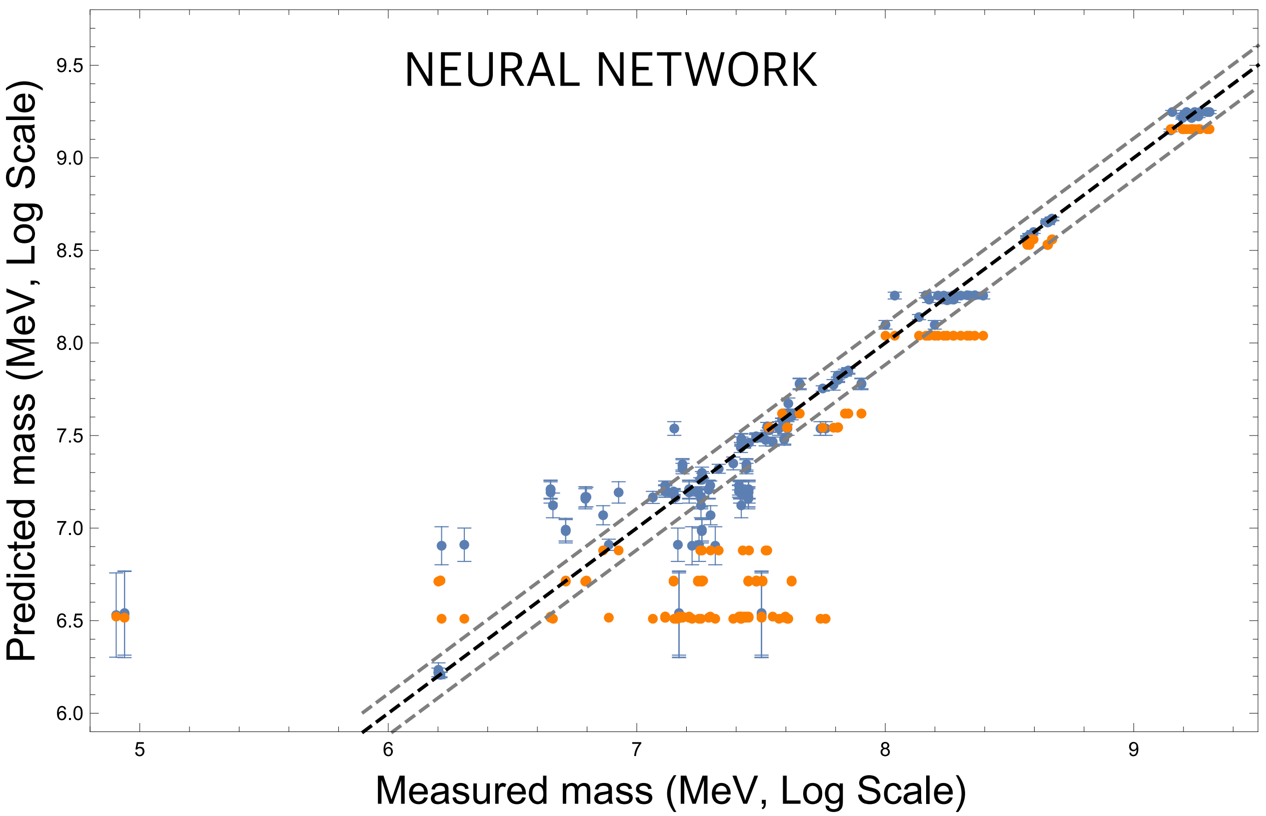


- We predict $$m = e^{\mu + \sigma^2/2}$$ from log normal distribution
  $$\sigma_m^2 = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

- Use 1000 runs

Constituent Quark Model

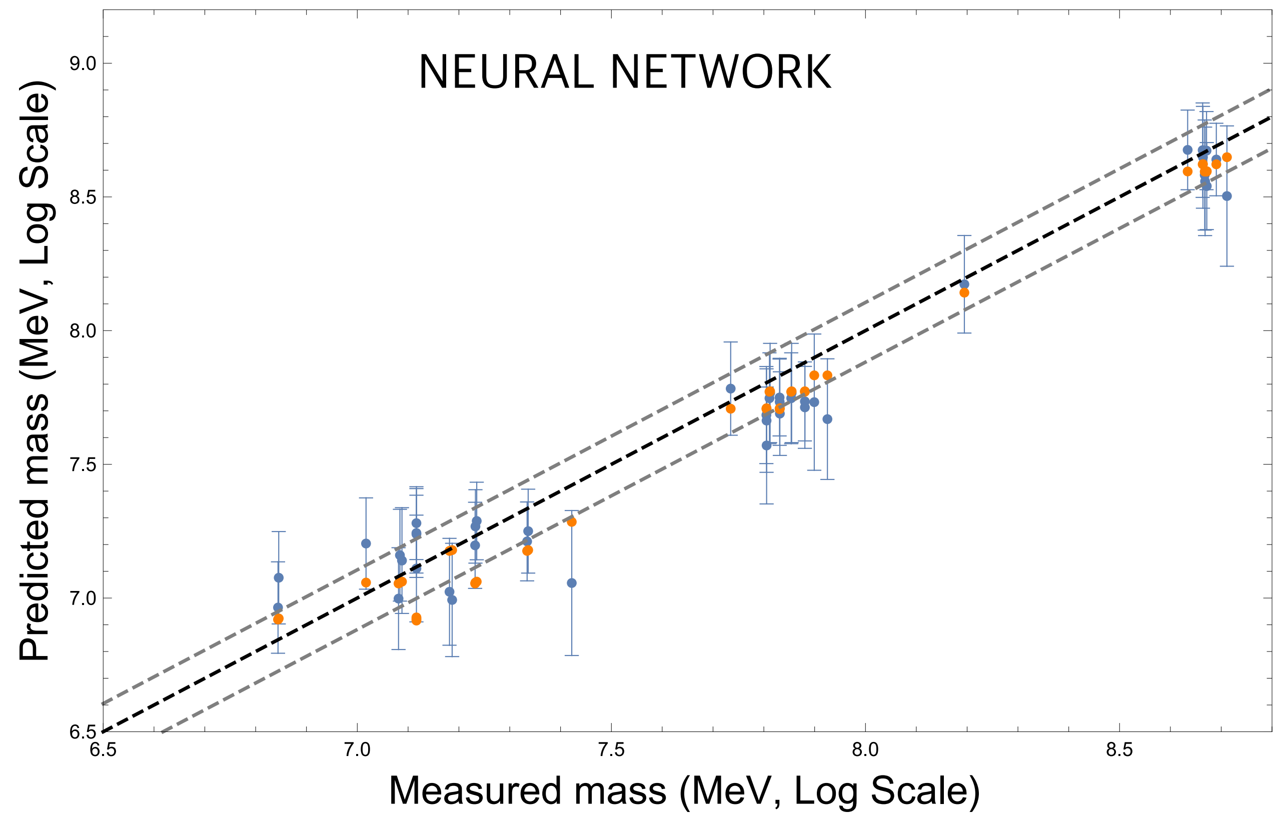
Mesons



13.0% error

39.3% error

Baryons



9.7% error

8.7% error

u : 336 MeV

d : 340 MeV

s : 486 MeV

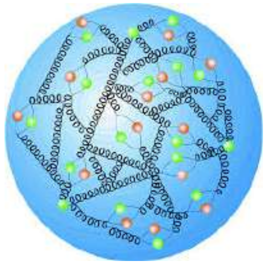
c : 1550 MeV

b : 4730 MeV

Constituent quark mass = QCD binding energy

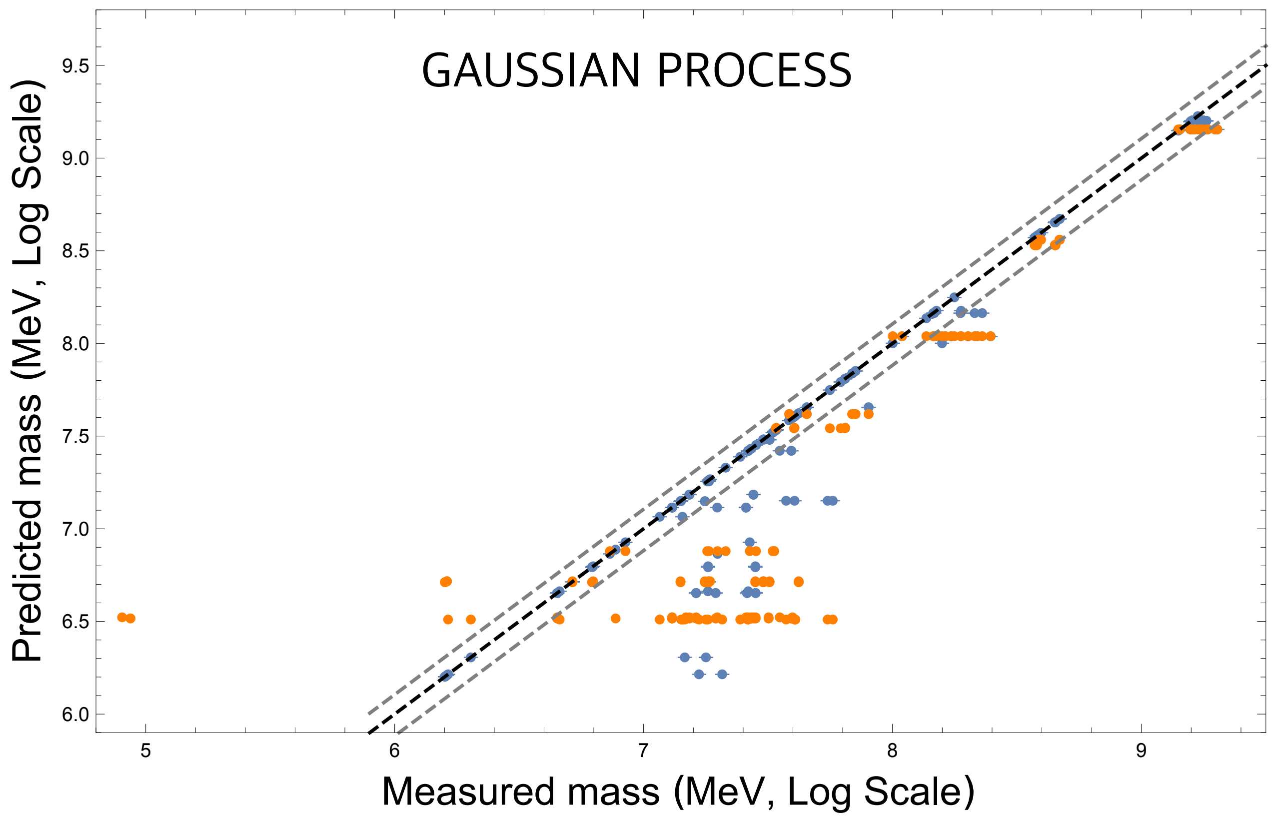
= amount of energy to add to spontaneously emit meson containing given valence quark

The neural network is doing something else!

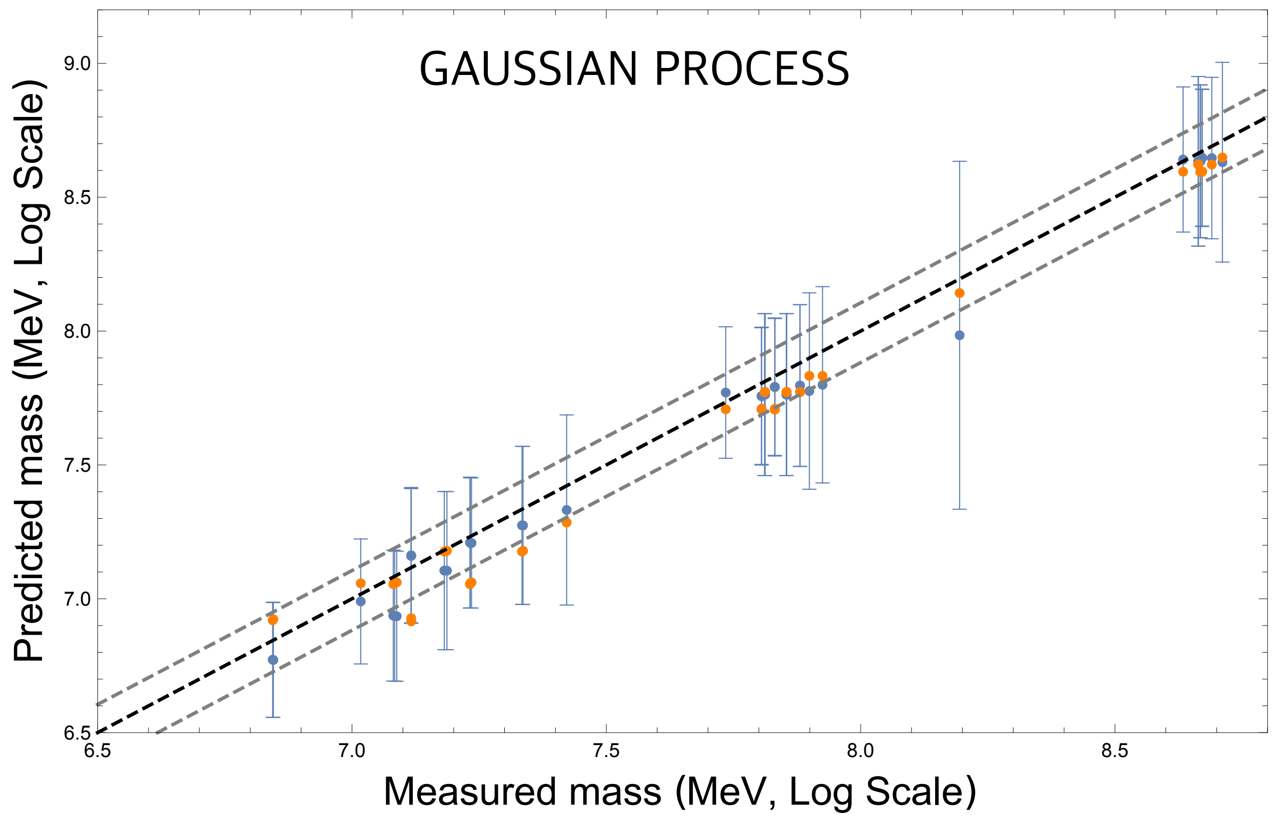


Gaussian Process

Mesons



Baryons



Mesons: 13.5% error; 39.3% error

Baryons: 3.4% error; 8.7% error

- Train with 14-vectors — augment $\vec{v}$ with rank to account for degenerate inputs
- Use $k_{\mathrm{RQ}}(x_i, x_j)$ kernel — 17 hyperparameters

Light Baryons

	p	n	Λ^0	Σ^+	Σ^0	Σ^-	Δ^{++}
#1	79.0	2.0	0.0	4.6	0.0	2.5	5.3
#2	13.9	26.7	0.0	12.5	1.2	5.5	22.5
#3	4.5	21.7	0.3	15.9	2.2	8.6	14.7
#4	1.9	14.7	0.6	17.3	5.7	8.5	13.6
#5	0.6	11.1	2.3	15.6	6.2	11.8	9.2
#6	0.1	9.1	3.9	9.6	8.0	10.7	5.8
#7	0.0	4.8	6.7	7.9	11.0	9.8	5.6

Light Baryons

	p	n	Λ^0	Σ^+	Σ^0	Σ^-	Δ^{++}
#1	79.0	2.0	0.0	4.6	0.0	2.5	5.3
#2	13.9	26.7	0.0	12.5	1.2	5.5	22.5
#3	4.5	21.7	0.3	15.9	2.2	8.6	14.7
#4	1.9	14.7	0.6	17.3	5.7	8.5	13.6
#5	0.6	11.1	2.3	15.6	6.2	11.8	9.2
#6	0.1	9.1	3.9	9.6	8.0	10.7	5.8
#7	0.0	4.8	6.7	7.9	11.0	9.8	5.6

$m_p = 1068 \pm 183$ MeV $\quad$ $m_p = 938.28$ MeV

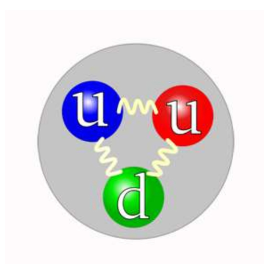
$m_n = 1205 \pm 206$ MeV $\quad$ $m_n = 939.57$ MeV

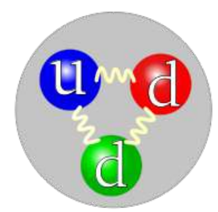
Gaussian process gives $m_p \approx m_n \approx 893 \pm 194$ MeV

CQM gives $m_p = 1012$ MeV , $m_n = 1016$ MeV

Light Baryons

$$\begin{pmatrix} |\pi^0\rangle \\ |\eta\rangle \\ |\eta'\rangle \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} |d\bar{d}\rangle \\ |u\bar{u}\rangle \\ |s\bar{s}\rangle \end{pmatrix}$$

- PDG's numbering scheme treats π^0 (lightest meson) as $d\bar{d}$
- Proton is 

and neutron is 

$$m_p = 1068 \pm 183 \text{ MeV} \qquad m_p = 938.28 \text{ MeV}$$

$$m_n = 1205 \pm 206 \text{ MeV} \qquad m_n = 939.57 \text{ MeV}$$

- Neural network learns this subtlety

Pentaquarks

$uudc\bar{c}$	$I\,(J^P)$	Measured Mass (MeV)	NN Pred (MeV)	GP Pred (MeV)	CQM Pred (MeV)
$\mathrm{P}_c(4312)^+$	$\frac{1}{2}\,(\frac{1}{2}^+)$	$4311.9 \pm 0.7^{+6.8}_{-0.6}$	$(4.2 \pm 1.2) \times 10^3$	3544 ± 923	
$\mathrm{P}_c(4440)^+$	$\frac{1}{2}\,(\frac{1}{2}^-)$	$4440.3 \pm 1.3^{+4.1}_{-4.7}$	$(4.1 \pm 1.1) \times 10^3$	3253 ± 846	4112
$\mathrm{P}_c(4457)^+$	$\frac{1}{2}\,(\frac{3}{2}^-)$	$4457.3 \pm 0.6^{+4.1}_{-1.7}$	$(4.5 \pm 1.1) \times 10^3$	3581 ± 932	

- Pentaquarks discovered by LHCb again with meson training

Tetraquarks?

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred (MeV)	GP Pred (MeV)	CQM Pred (MeV)
$a_0(980)$	$1\,(0^+)$	980 ± 20	$u\bar{u}$	1277 ± 246	511 ± 34	680
			$u\bar{u}s\bar{s}$ $(K\overline{K})$	2172 ± 466	1713 ± 68	1652
$f_0(980)$	$0\,(0^+)$	990 ± 20	$d\bar{d}$	921 ± 117	977 ± 37	672
			$d\bar{d}s\bar{s}$ $(K\overline{K})$	1592 ± 401	1132 ± 312	1644

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred (MeV)	GP Pred (MeV)	CQM Pred (MeV)
$D^*_{s0}(2317)^{\pm}$	$0\,(0^+)$	2317.8 ± 0.5	$c\bar{s}$	2640 ± 433	2434 ± 700	2036
			$c\bar{u}u\bar{s}$ (DK)	4326 ± 925	2474 ± 826	2858
$D_{s1}(2460)^{\pm}$	$0\,(1^+)$	2459.5 ± 0.6	$c\bar{s}$	2547 ± 39	2535 ± 0.003	2036
			$c\bar{u}u\bar{s}$ (D^*K)	3431 ± 544	2560 ± 788	2858

- Discriminate between composition hypotheses for new QCD resonances
- Tetraquarks are disfavored by both neural network and Gaussian process

Semi-Empirical Mass Formula

- Mass of atomic nucleus with $A = Z + N$

$$m(Z,N) = Zm_p + Nm_n - E_B(Z,N)$$

$$E_B(Z,N) = a_V A - a_S A^{\frac{2}{3}} - a_C \frac{Z(Z-1)}{A^{\frac{1}{3}}} - a_A \frac{(A-2Z)^2}{A} - \delta(Z,N)$$

volume surface Coulomb asymmetry pairing

$$\delta(Z,N) = \begin{cases} +\delta_0 & Z, N \text{ even} \\ -\delta_0 & Z, N \text{ odd} \\ 0 & \text{otherwise} \end{cases} \quad , \quad \delta_0 = a_P A^{-\frac{1}{2}}$$

Weizsäcker (1935)

- Does such a formula exist for baryons?

- Machine learning experiments suggest there is an $m(\vec{v})$ formula

Semi-Empirical Mass Formula

- In large-N limit, interaction between quarks given by Hartree–Fock potential

  Witten (1979)

  Density grows as number of particles increases

- In nucleus, density does not increase with number of particles

  In fact, binding energy term negligible as Z, N get large

- In baryon case, interaction energy grows with N

- The formula is therefore more complicated than in liquid drop model

Universal Approximation Theorem

- Suppose some more complicated formula exists

- Feedforward neural network, sigmoid activation function, single hidden layer with finite number of neurons can approximate continuous functions on compact subsets of $\mathbb{R}^n$

Cybenko (1989)
Hornyk (1991)

- Surprise here is simplicity: 50 neurons, 196 meson dataset for training

- Principal component analysis indicates $c, \ \bar{c}, \ b, \ \bar{b}$ count most, then I

- Antibaryon masses to within 6% of baryon masses; so machine learning figures out antiparticles on its own

Baryons as Solitons

- Mesons have mass that scales like $\dfrac{1}{N}$

- In real world QCD, $\dfrac{\#}{N^2}$ is a small parameter; perturbation theory works

- Leads to approximate decoupling of degrees of freedom in S-matrix picture

Witten (1979)

- Machine learning baryon spectrum from mesons is consistent with this picture

Prospectus

- Glueballs
- Learn Regge trajectories
- Towards QCD string
- Gell-Mann–Okubo mass formula
- Better chiral perturbation theory
- Faster than lattice QCD, but certainly doesn't supplant it

Collaborators

Yarin Gal

Damián Kaloni Mayorga Peña

Challenger Mishra

arXiv:2003.10445

Challenge

- How does a black box learn semantics without knowing syntax?

— Generally unpublished failed experiments indicate what doesn't work

— Knowing that there are approximate functions $f(\vec{J})$ and $m(\vec{v})$

can we find analytic expressions by opening the black box?

Challenge

- How does a black box learn semantics without knowing syntax?

  — Generally unpublished failed experiments indicate what doesn't work

  — Knowing that there are approximate functions $f(\vec{J})$ and $m(\vec{v})$ can we find analytic expressions by opening the black box?

- Can artificial intelligence do interesting research?

  — *cf.* new jōseki in go — AlphaGo Zero (2017)

  — Proofs in real analysis — Ganesalingam, Gowers (2013)

  — Proof assistants — Voevodsky (2014)

hep-th

- Use machine learning to classify papers into **arXiv** categories
- 65% success at exact subject, 87% success at formal vs. phenomenology
- Mapping words to vectors contextually, we discover syntactic identities

$$\text{Paris} - \text{France} + \text{Italy} = \text{Rome}$$

$$\text{king} - \text{man} + \text{woman} = \text{queen}$$


He, VJ, Nelson (2018)

hep-th

- Use machine learning to classify papers into **arXiv** categories
- 65% success at exact subject, 87% success at formal vs. phenomenology
- Mapping words to vectors contextually, we discover syntactic identities

$$\text{Paris} - \text{France} + \text{Italy} = \text{Rome}$$

$$\text{king} - \text{man} + \text{woman} = \text{queen}$$

- An idea generating machine for **hep-th**:

$$\text{symmetry} + \text{black hole} = \text{Killing}$$

$$\text{symmetry} + \text{algebra} = \text{group}$$

$$\text{black hole} + \text{QCD} = \text{plasma}$$

$$\text{spacetime} + \text{inflation} = \text{cosmological constant}$$

$$\text{string theory} + \text{Calabi}-\text{Yau} = \text{M}-\text{theory} + G_2$$

He, VJ, Nelson (2018)

Thank You!