Machine Learning as a Discovery Tool

Vishnu Jejjala
University of the Witwatersrand

Accelerate Science Winter School
Cambridge University
4 February 2021

The Beginning

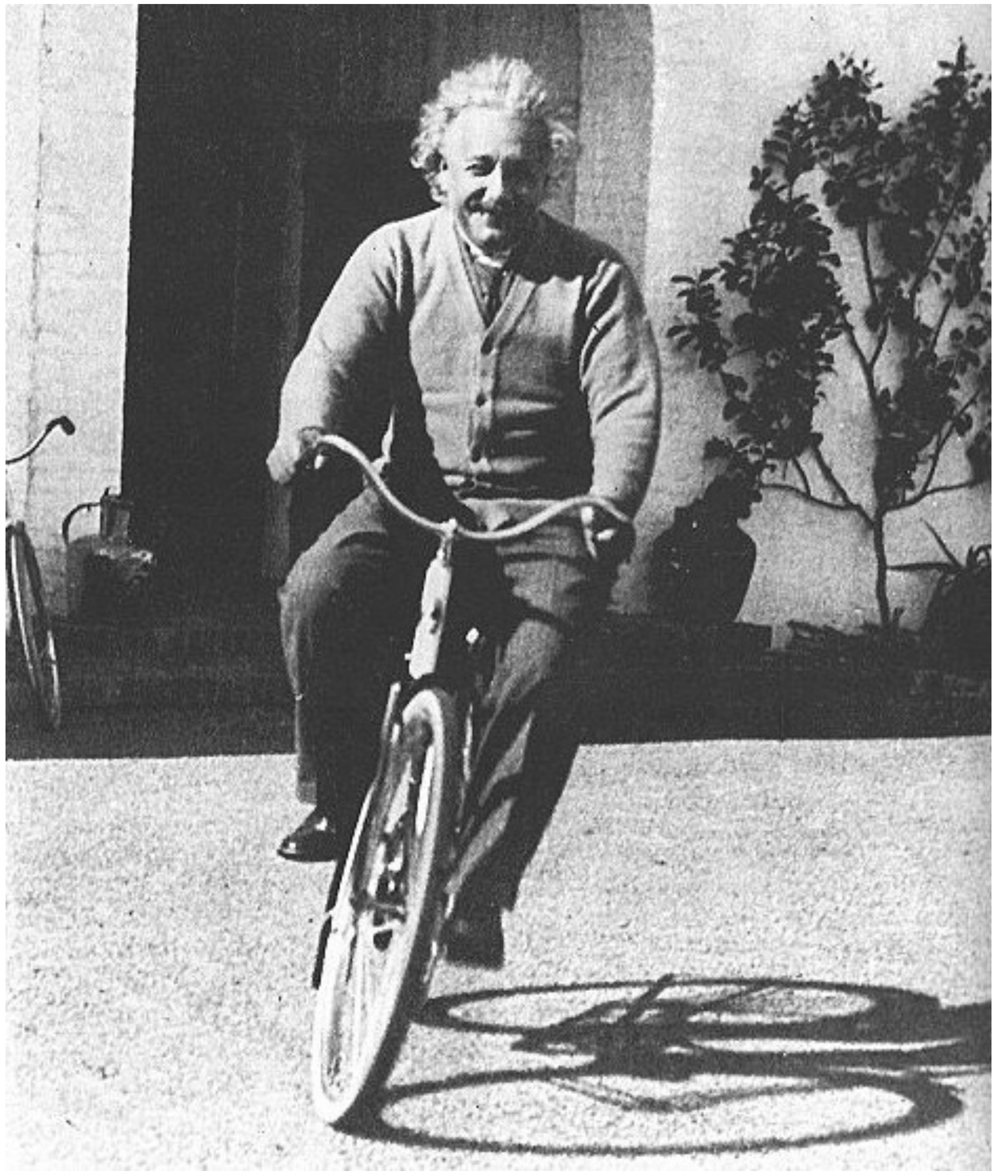
Image: Karen Carr

Fundamental Questions

- How did the Universe begin?
- Why is the world the way it is?
- Could it have been some other way?
- What is the fundamental explanation for space, time, and matter?

How We Do Physics

- Interrogate a theory at its limits and test it against other theories
- Investigate the tensions



A gedankenexperiment

Turn on the headlight of your bicycle

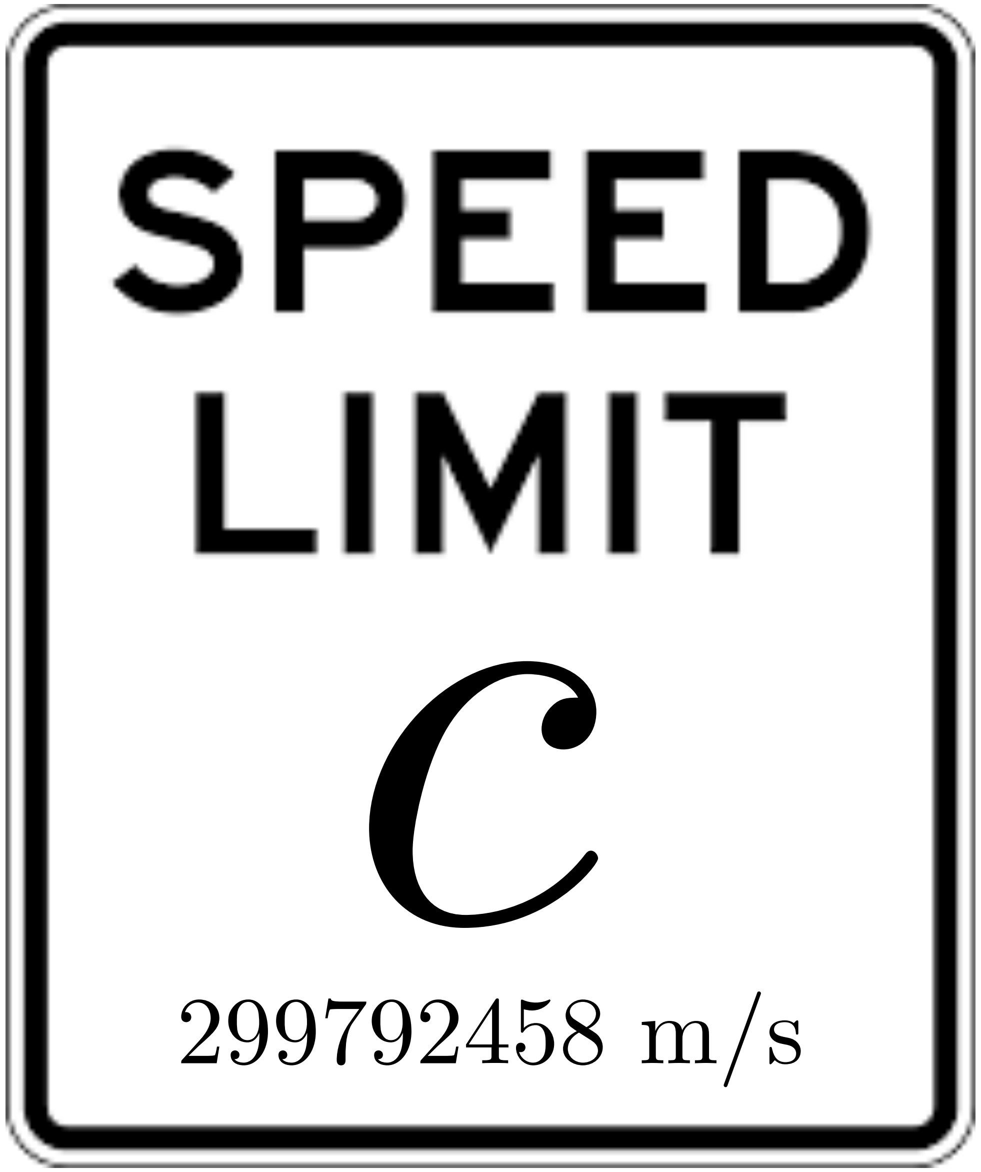
Suppose you bicycle faster than light

What do you see?

This thought experiment brings Galileo and Maxwell into tension

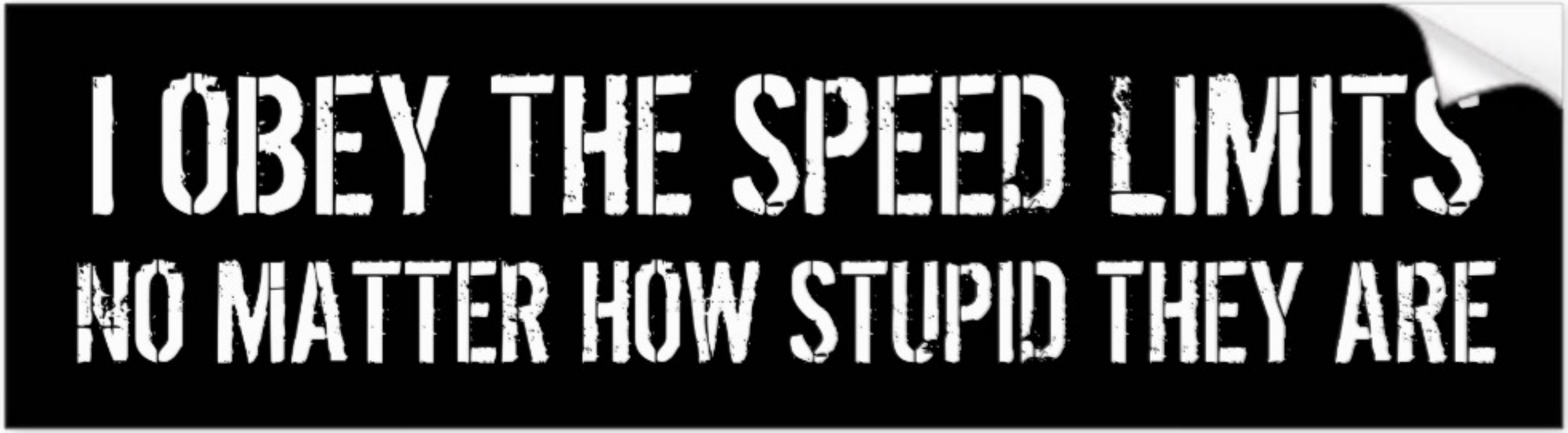
Special Relativity

- Every observer measures the same speed of light
- The Universe has a speed limit



Special Relativity

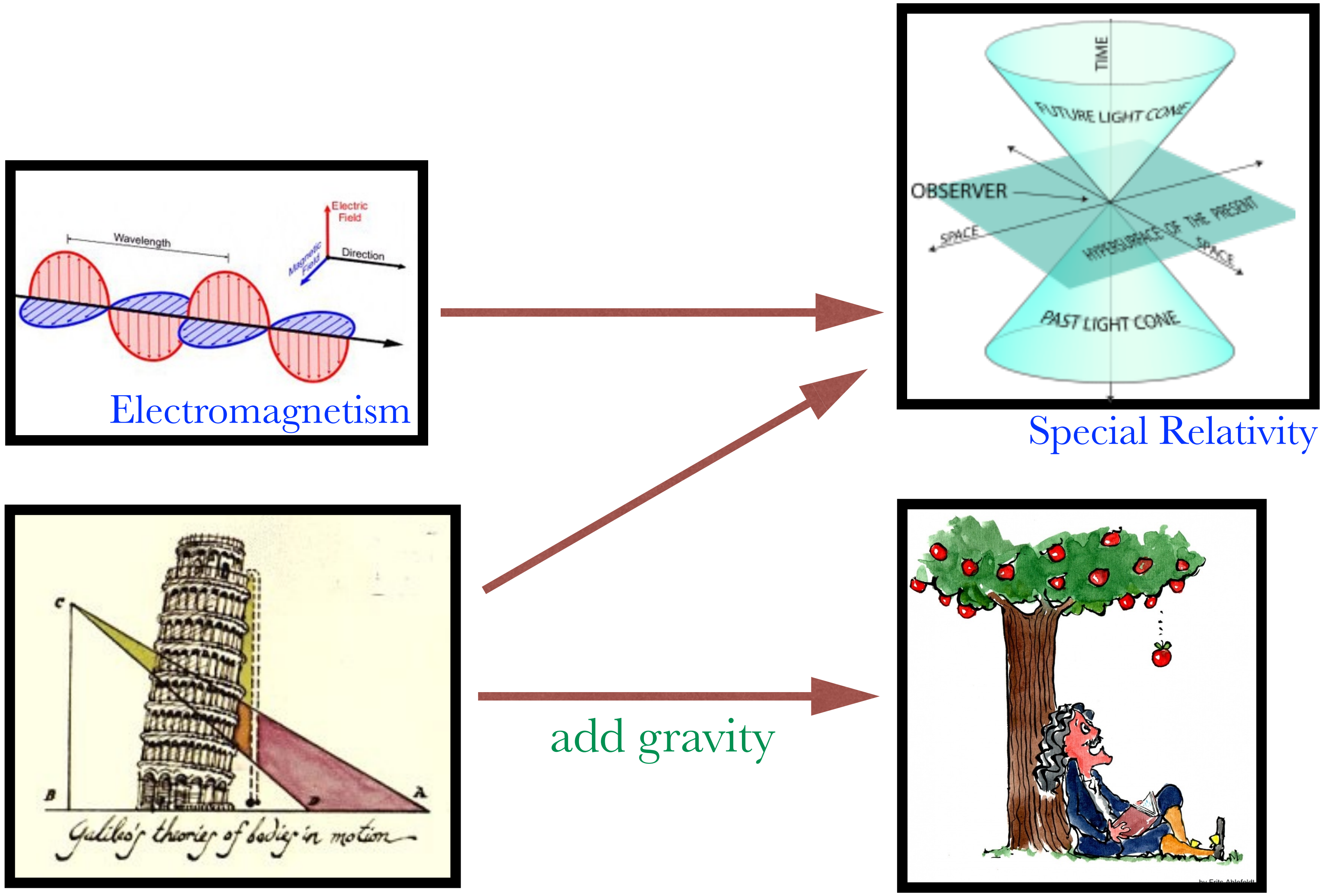
- Every observer measures the same speed of light
- The Universe has a speed limit



I OBEY THE SPEED LIMITS
NO MATTER HOW STUPID THEY ARE

Theories Beget Theories

- By testing electromagnetism against Galilean mechanics, we arrive at the special theory of relativity



Electromagnetism

Special Relativity

add gravity

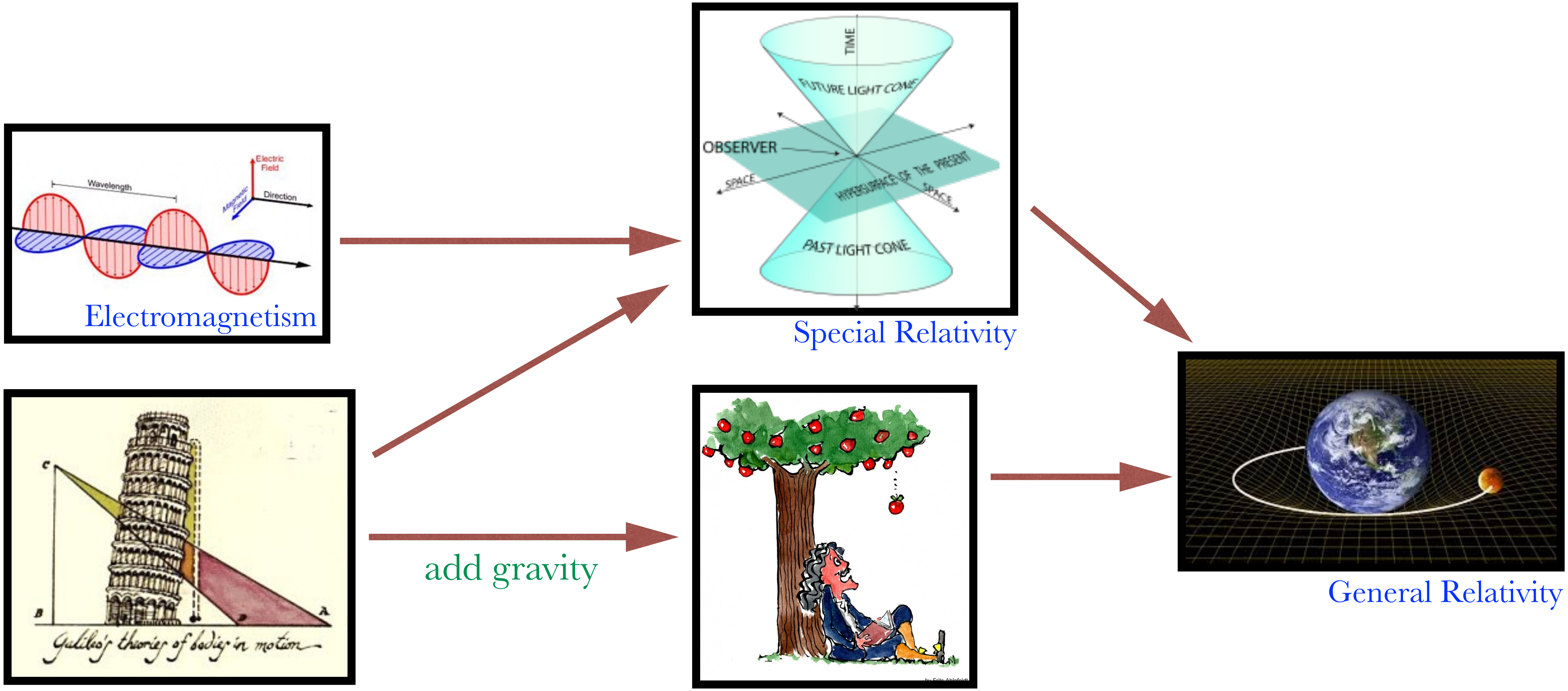
$$\vec{\mathbf{F}} = m\vec{\mathbf{a}}$$

$$\vec{\mathbf{F}} = -\frac{G_N M m}{r^2}\hat{\mathbf{r}}$$

Theories Beget Theories

- By testing electromagnetism against Galilean mechanics, we arrive at the special theory of relativity

- Let’s continue on this path



Electromagnetism

Galilean Mechanics

add gravity

Special Relativity

Newtonian Mechanics

General Relativity

General Relativity

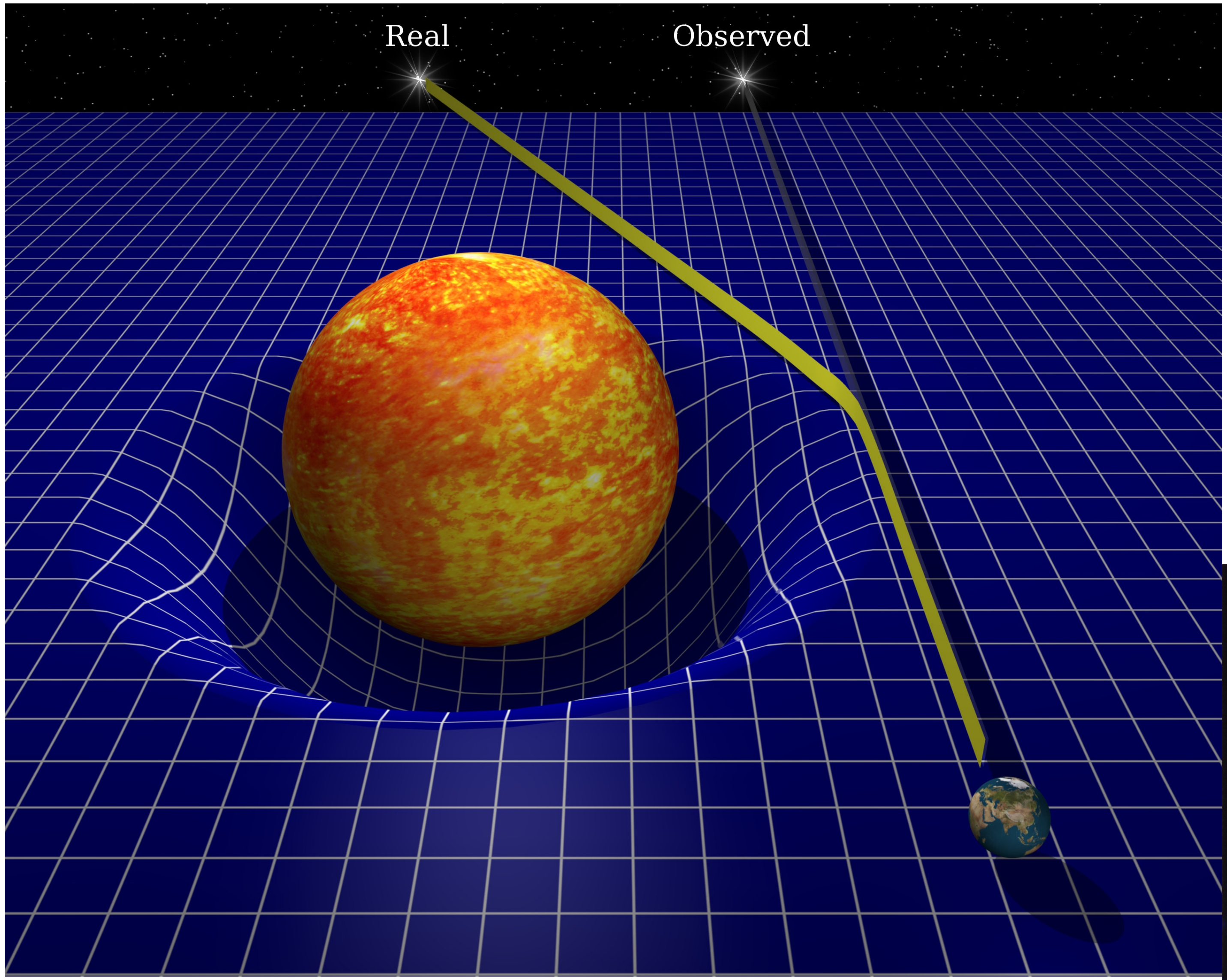


Image: Dave Jarvis

Another gedankenexperiment

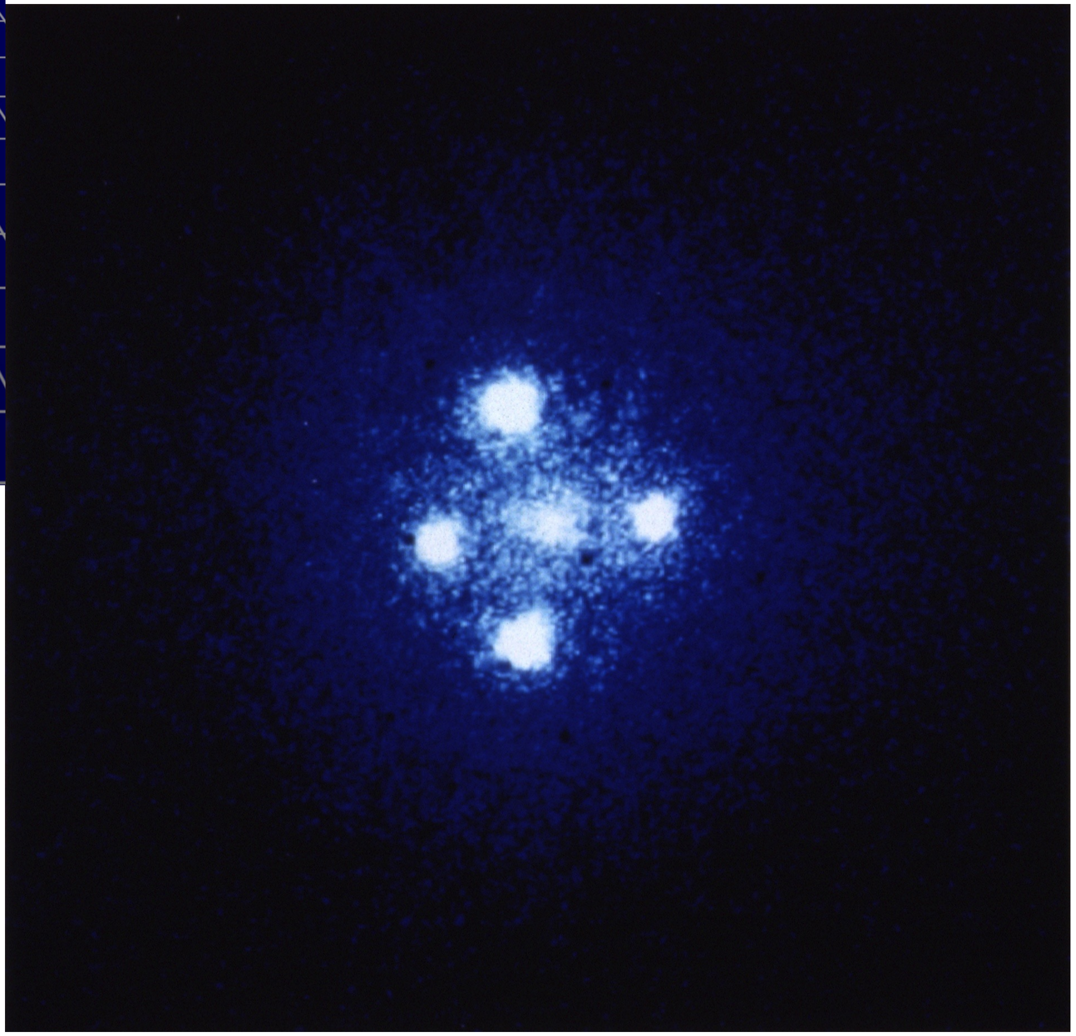
What happens if the Sun suddenly disappeared?

Tension between Newton and Einstein

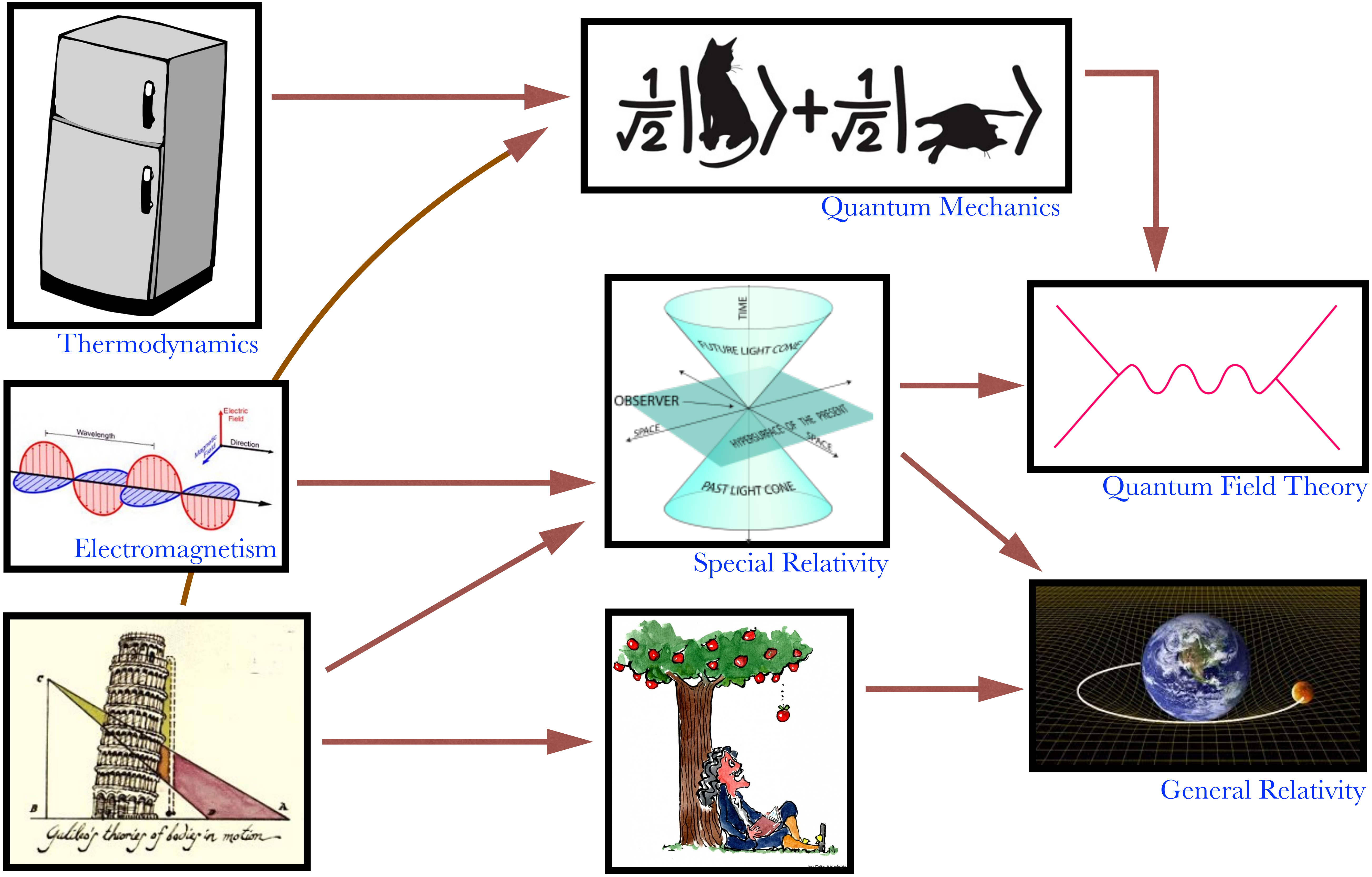
Force of gravity is geometry

Verified from microns to cosmic scales

Image: ESA/NASA (HST)



Theory Space



Thermodynamics

Quantum Mechanics

Electromagnetism

Special Relativity

Quantum Field Theory

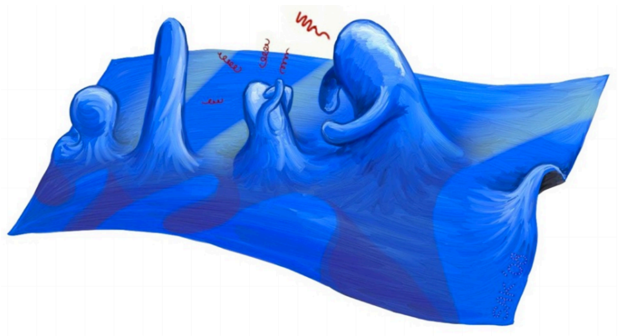
General Relativity

Galilean Mechanics

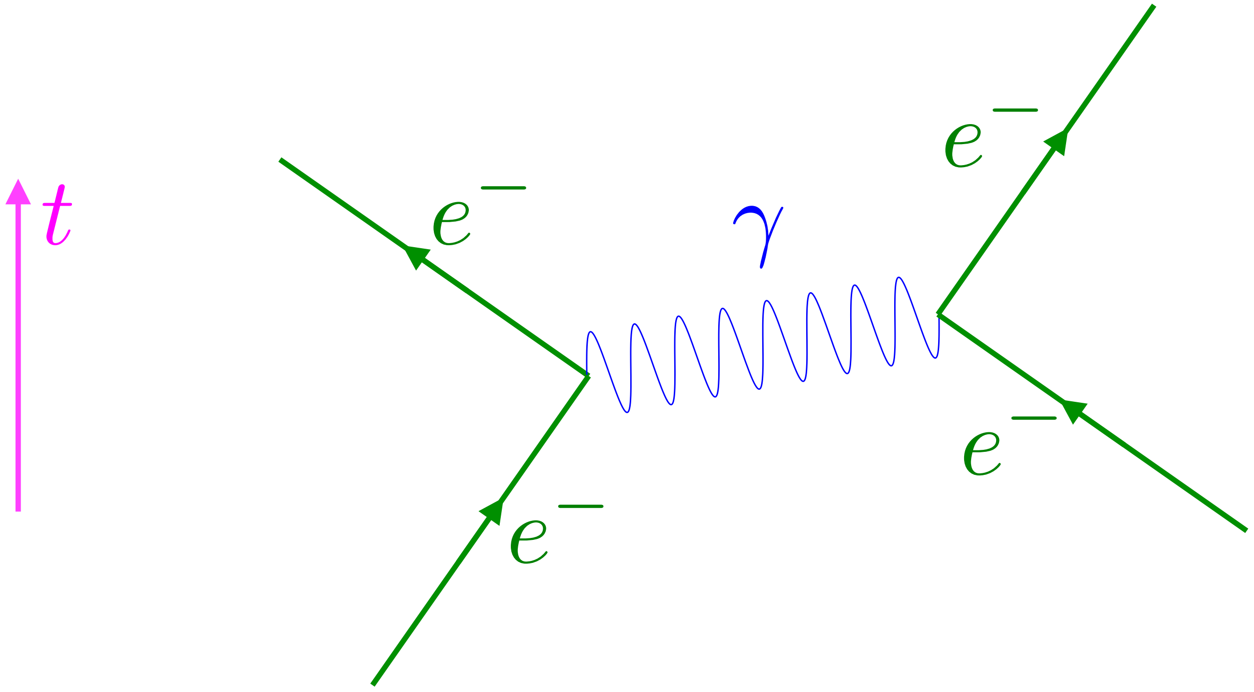
Newtonian Mechanics

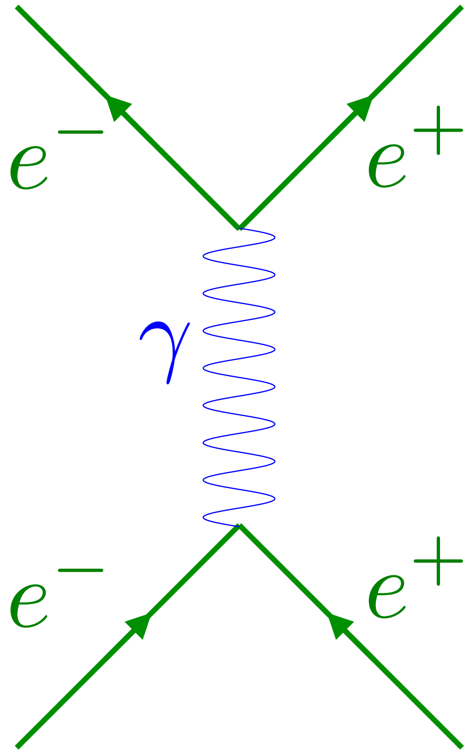
Quantum Field Theory

- 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





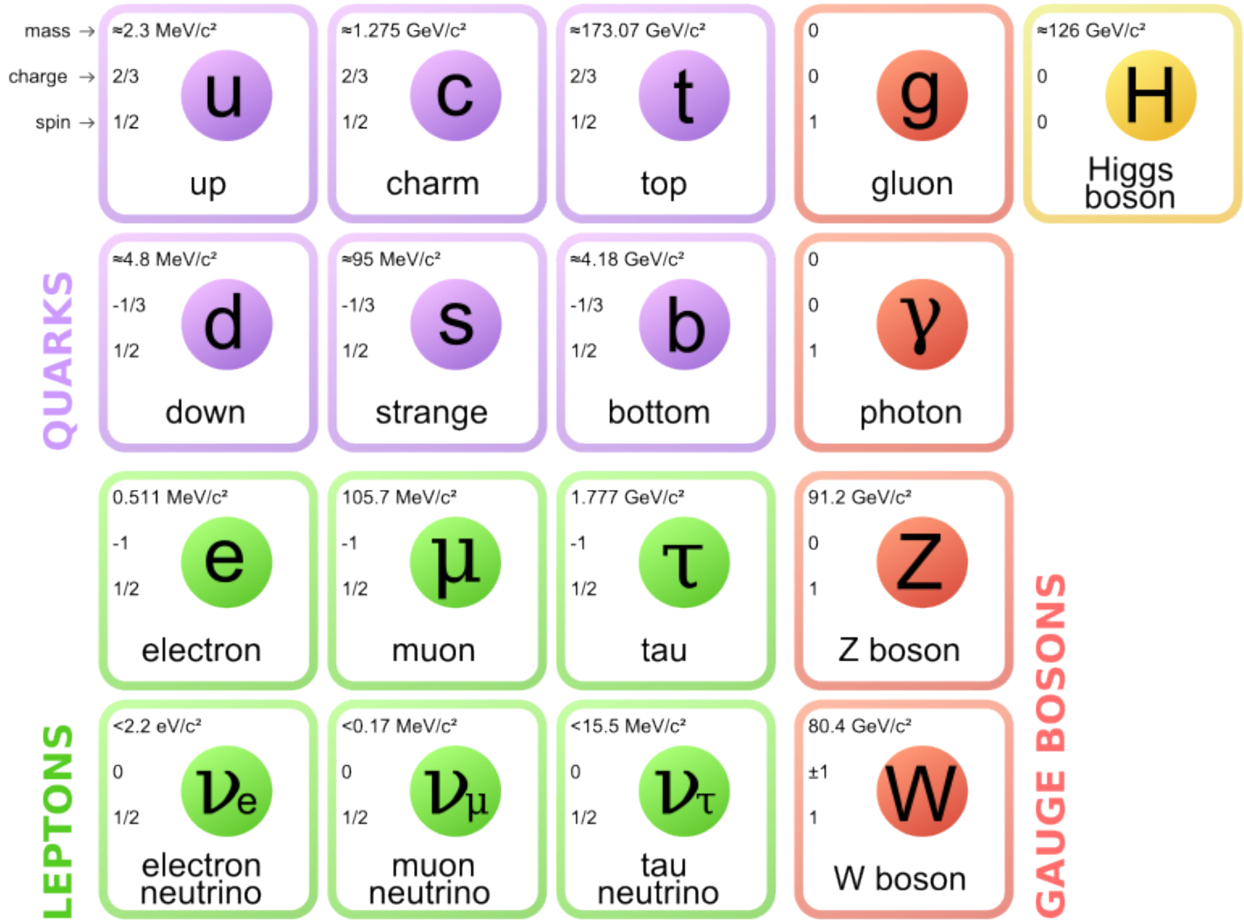
- Electrons and positrons interact by exchanging photons, for example

Quantum Field Theory

- Fundamental forces are described by quantum field theory

- Standard Model

electromagnetism
weak force
strong force
Higgs effect

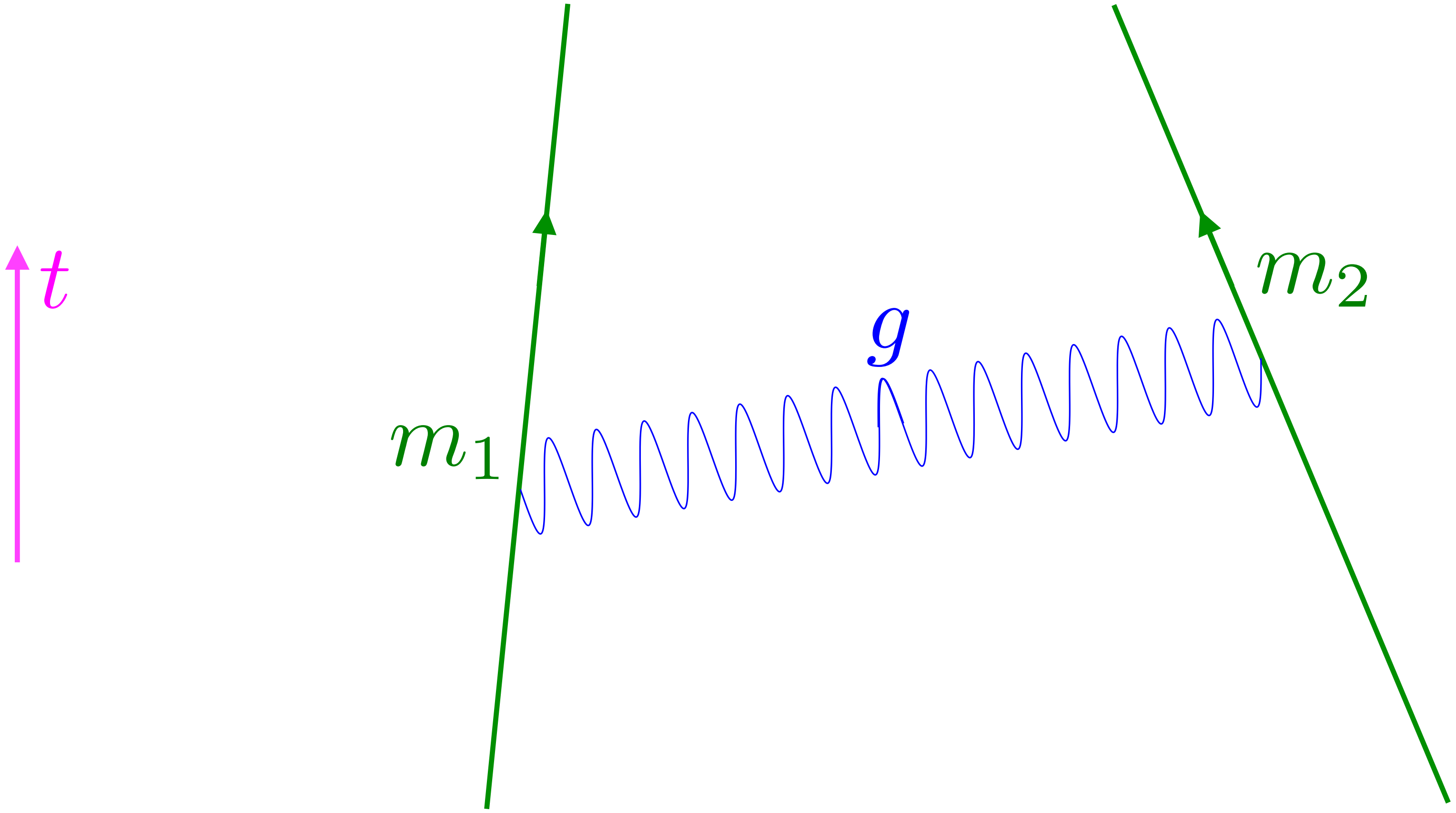


$$\alpha_{\rm exp}^{-1} = 137.035999139(31)$$

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

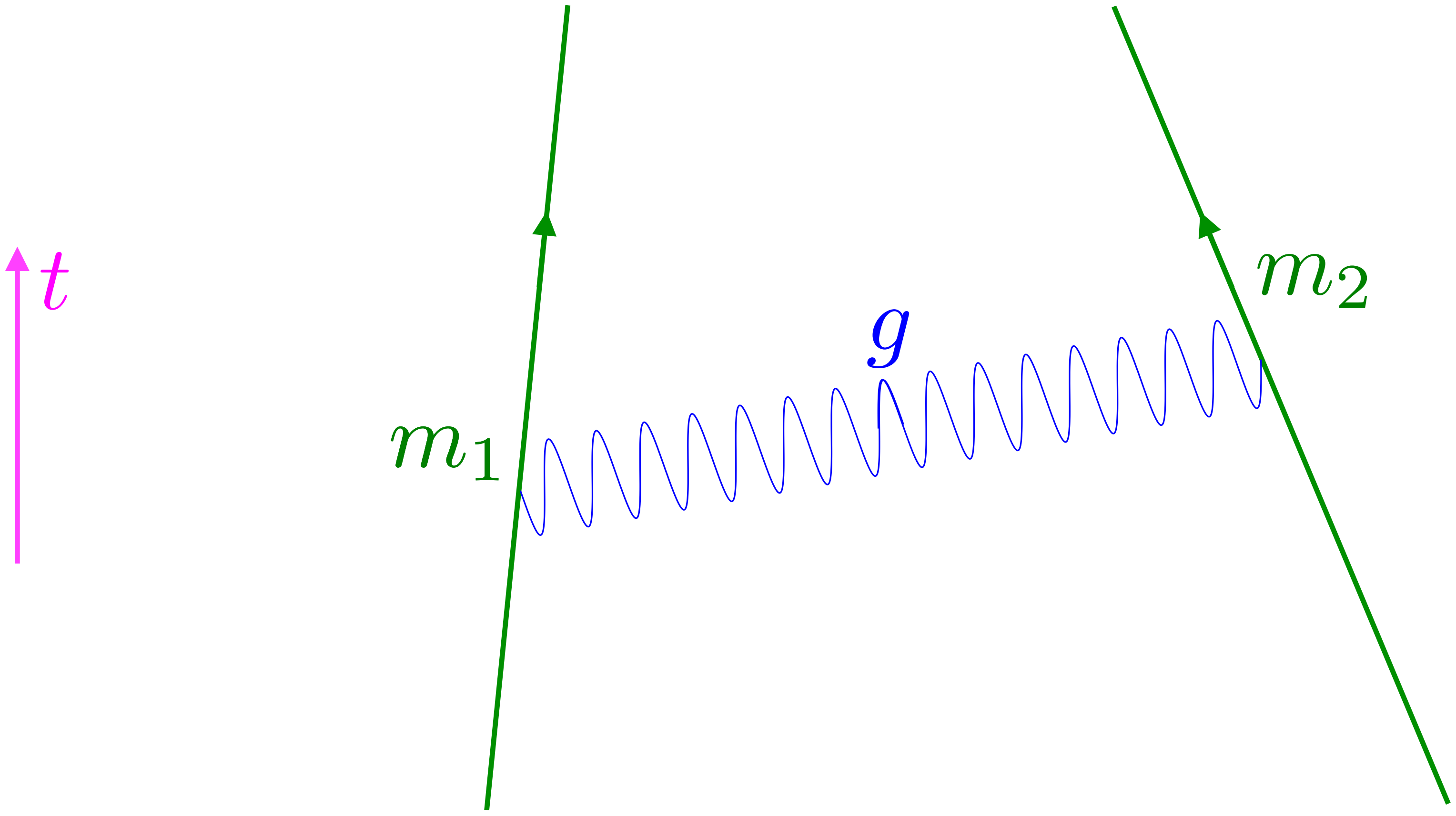
One Theory of Physics

- Gravity is a response to curvature, but we experience this as a force
- Matter couples to geometry via mass
- What happens if we treat geometry as a quantum field?



One Theory of Physics

- Gravity is a response to curvature, but we experience this as a force
- Matter couples to geometry via mass
- What happens if we treat geometry as a quantum field?



$$+\ldots \sim \infty$$

What Went Wrong?

- General relativity explains the dynamical response of geometry to the presence of matter or energy and conversely the dynamical response of matter to the curvature of spacetime

- In a quantum Universe, things fluctuate due to the uncertainty principle

$$\boxed{\sigma_x \, \sigma_p \geq \frac{\hbar}{2}}$$

- Because spacetime itself fluctuates at the quantum level, one of the central assumptions of general relativity, that geometry is smooth, breaks down

- Quantum field theory is not the organizing principle

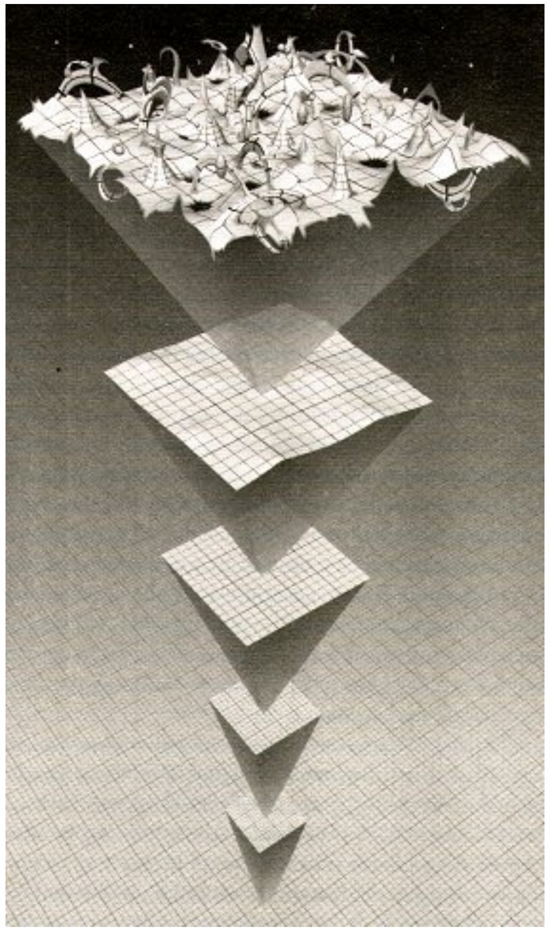
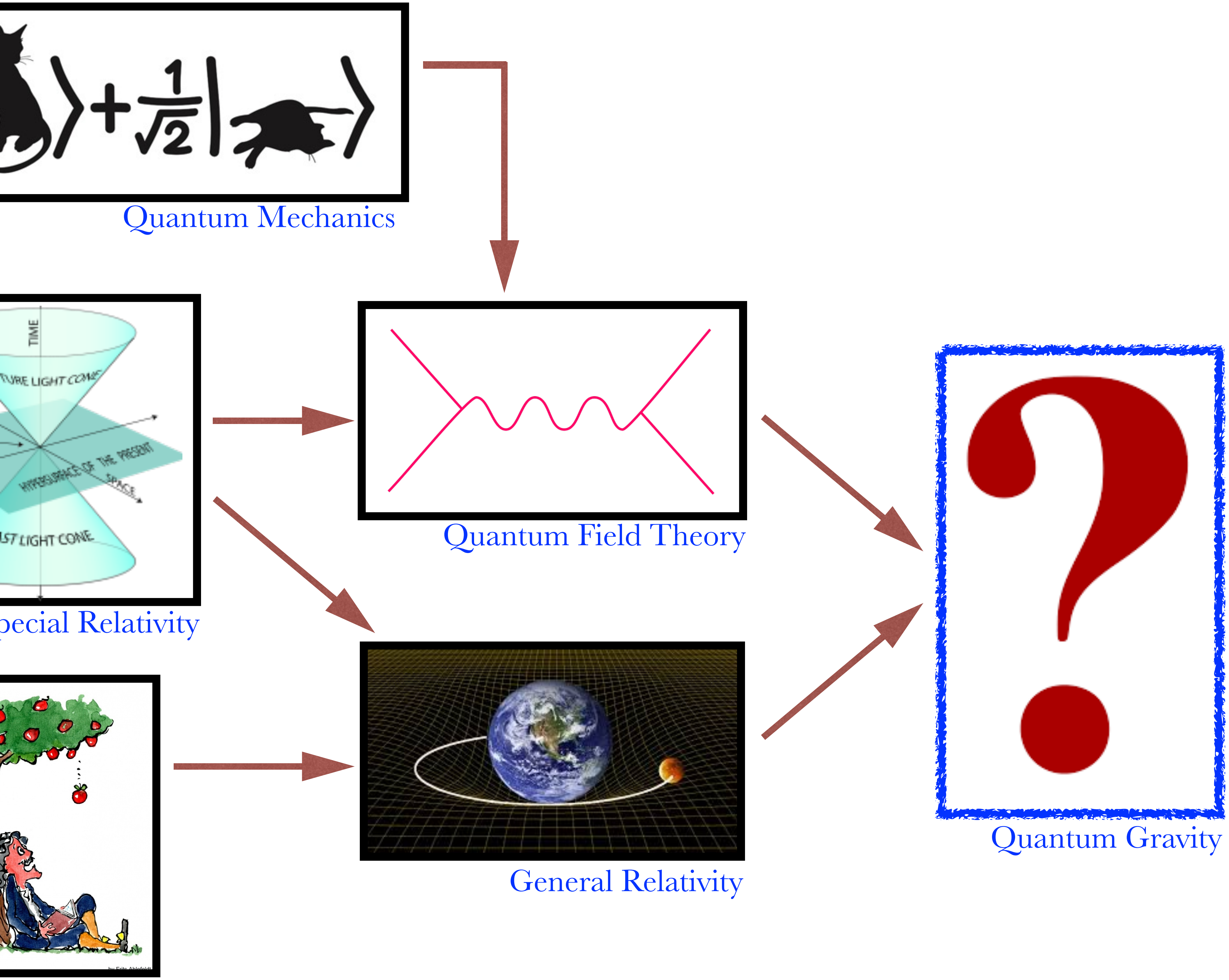


Image: Brian Greene

A New Hope



Quantum Mechanics

pecial Relativity

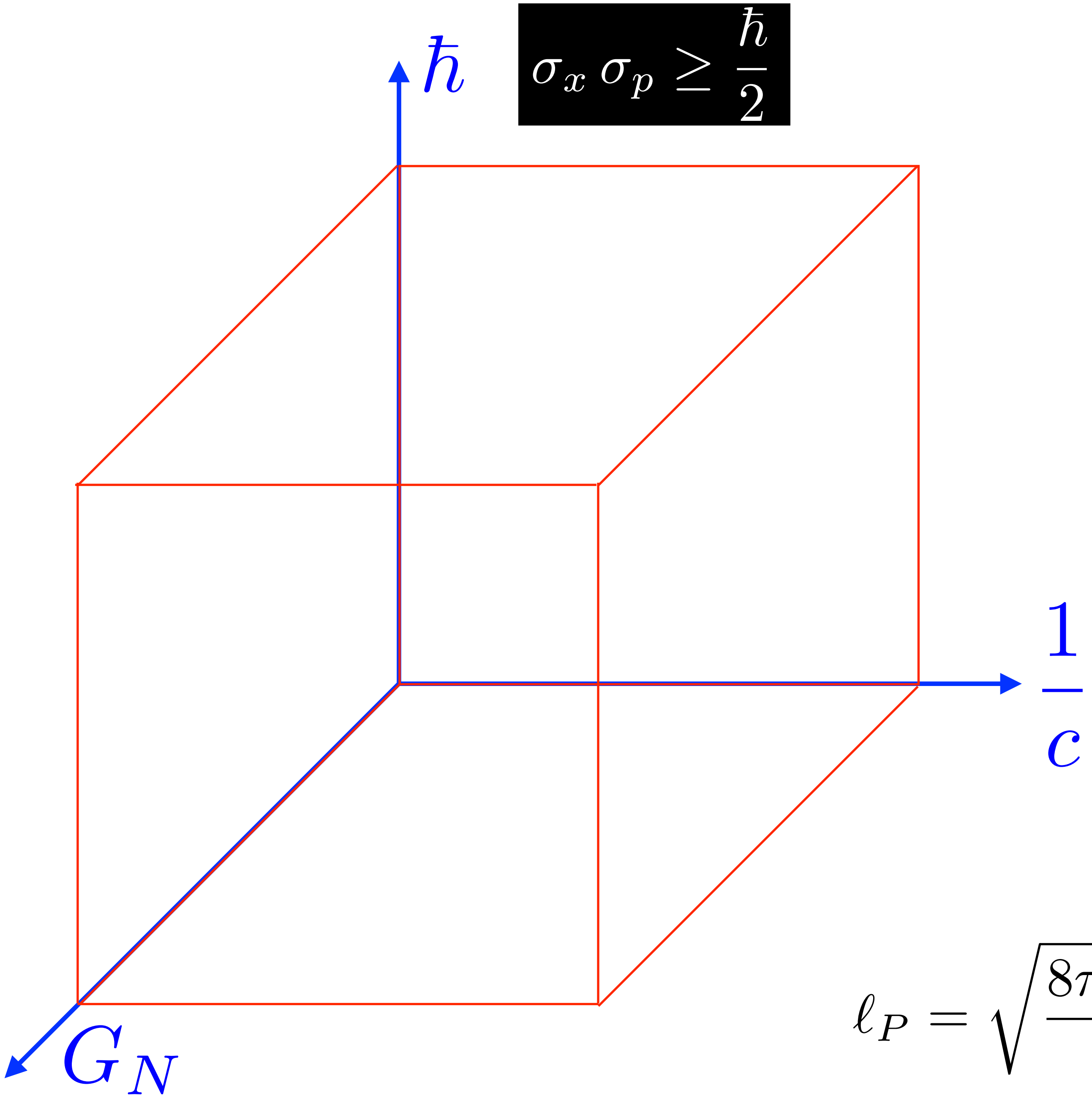
n Mechanics

Quantum Field Theory

General Relativity

Quantum Gravity

Bronstein's Cube



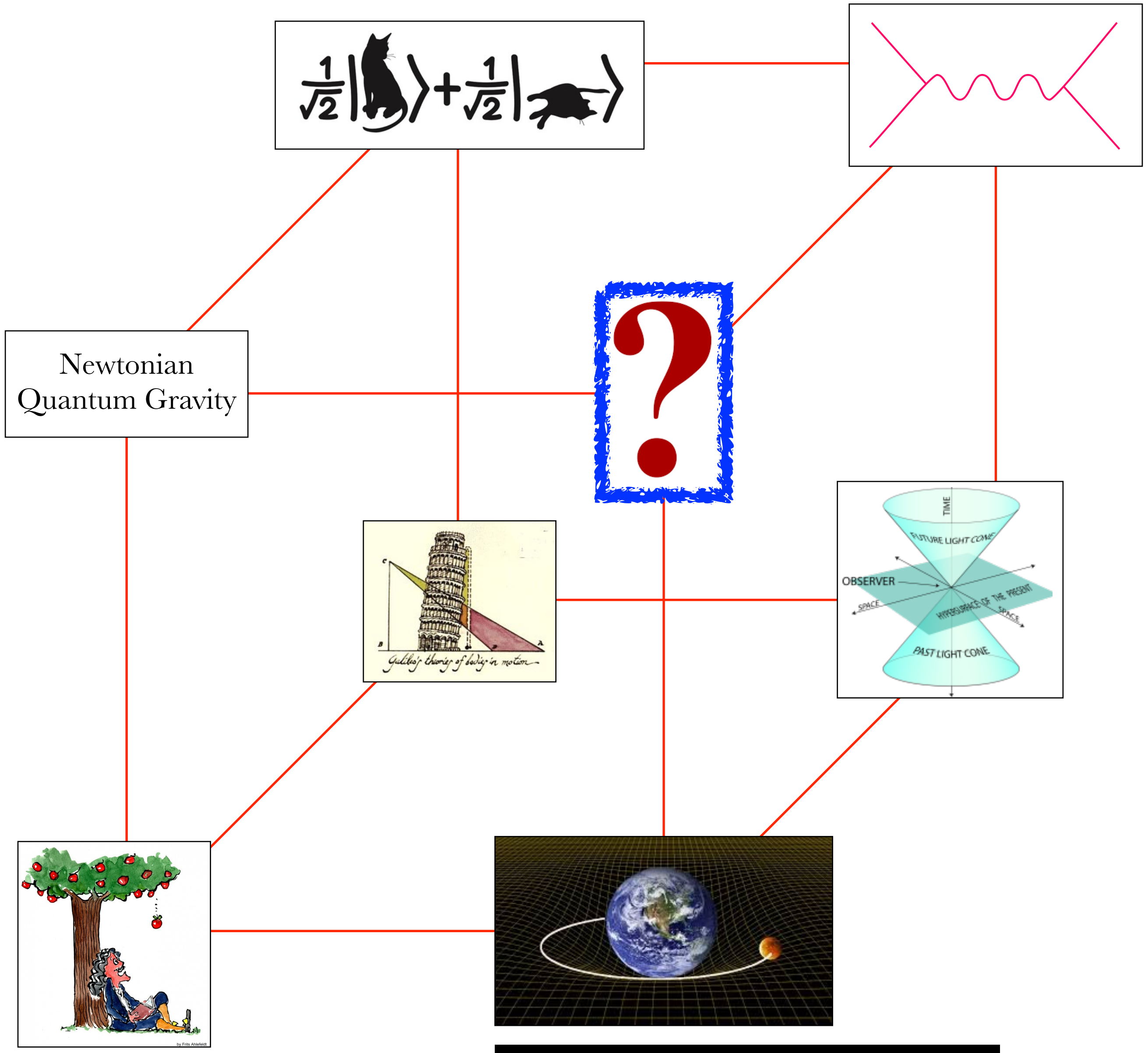
$$\sigma_x \, \sigma_p \geq \frac{\hbar}{2}$$

I OBEY THE SPEED LIMITS
NO MATTER HOW STUPID THEY ARE

MASSES ATTRACT

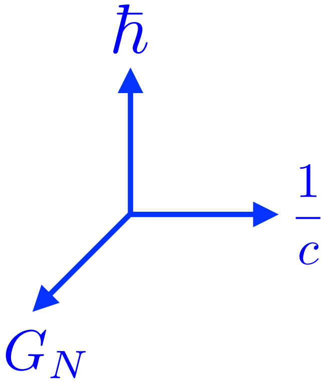
$$\ell_P = \sqrt{\frac{8\pi G_N \hbar}{c^3}} \approx 8.1 \times 10^{-35} \text{ m}$$

Bronstein's Cube



FIELDS INTERACT
VIA GAUGE FORCES

PARTICLES ARE
FIELD EXCITATIONS



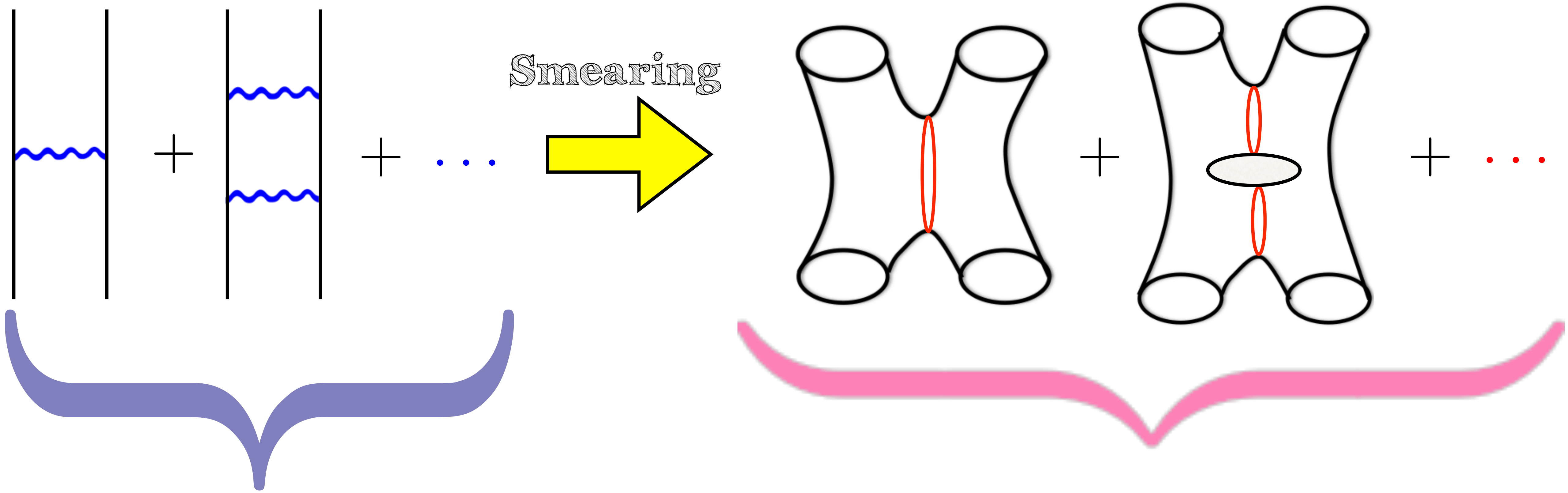
ENERGY COUPLES
TO GEOMETRY

SPACETIME IS
DYNAMICAL

String Theory

Gravity as a QFT

Gravity from String Theory



These are infinite

These are four dimensional

These are finite

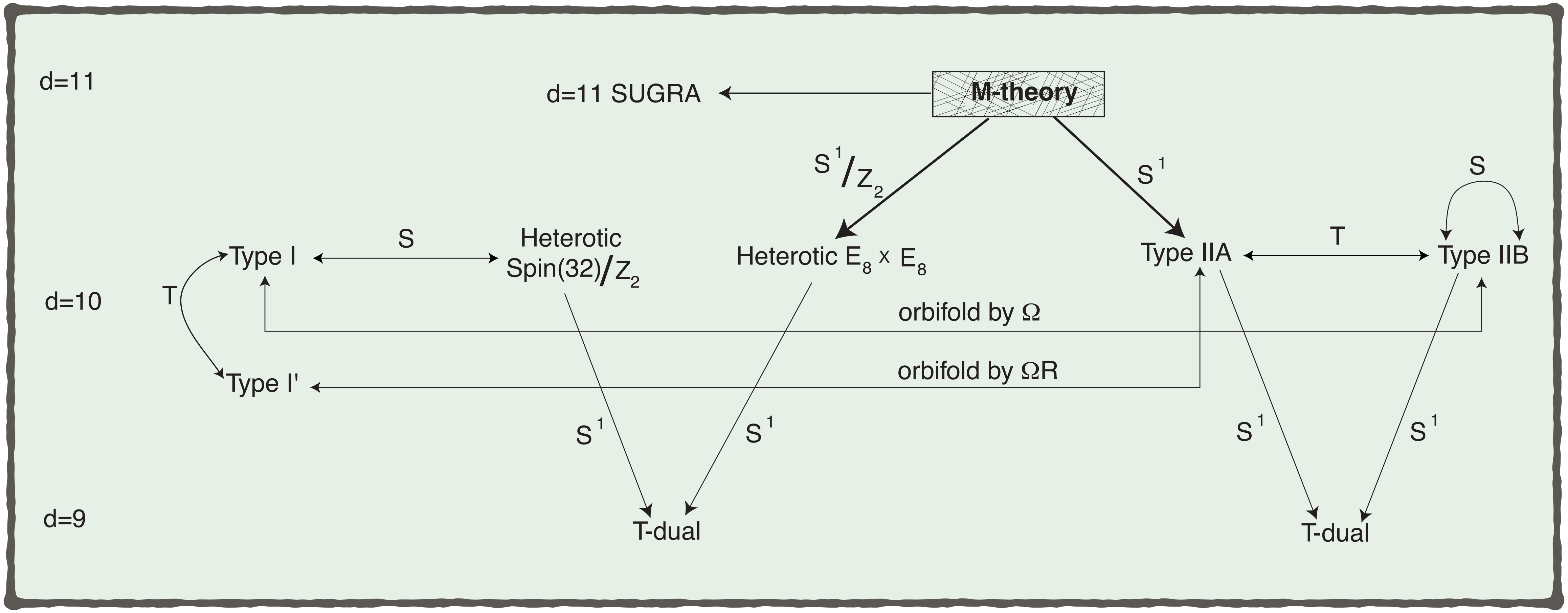
These are ten dimensional

[To prove the consistency of string theory we use the remarkable fact that $\sum_{n=1}^{\infty} n \longrightarrow -\frac{1}{12}$]

- $X^\mu : \Sigma \to \mathcal{M}$

  Sigma model on the string worldsheet gives general relativity

String Theory



- String theory is in fact a web of interconnected theories in ten (or eleven or twelve) dimensions

- We experience only four dimensions. So how do we proceed?

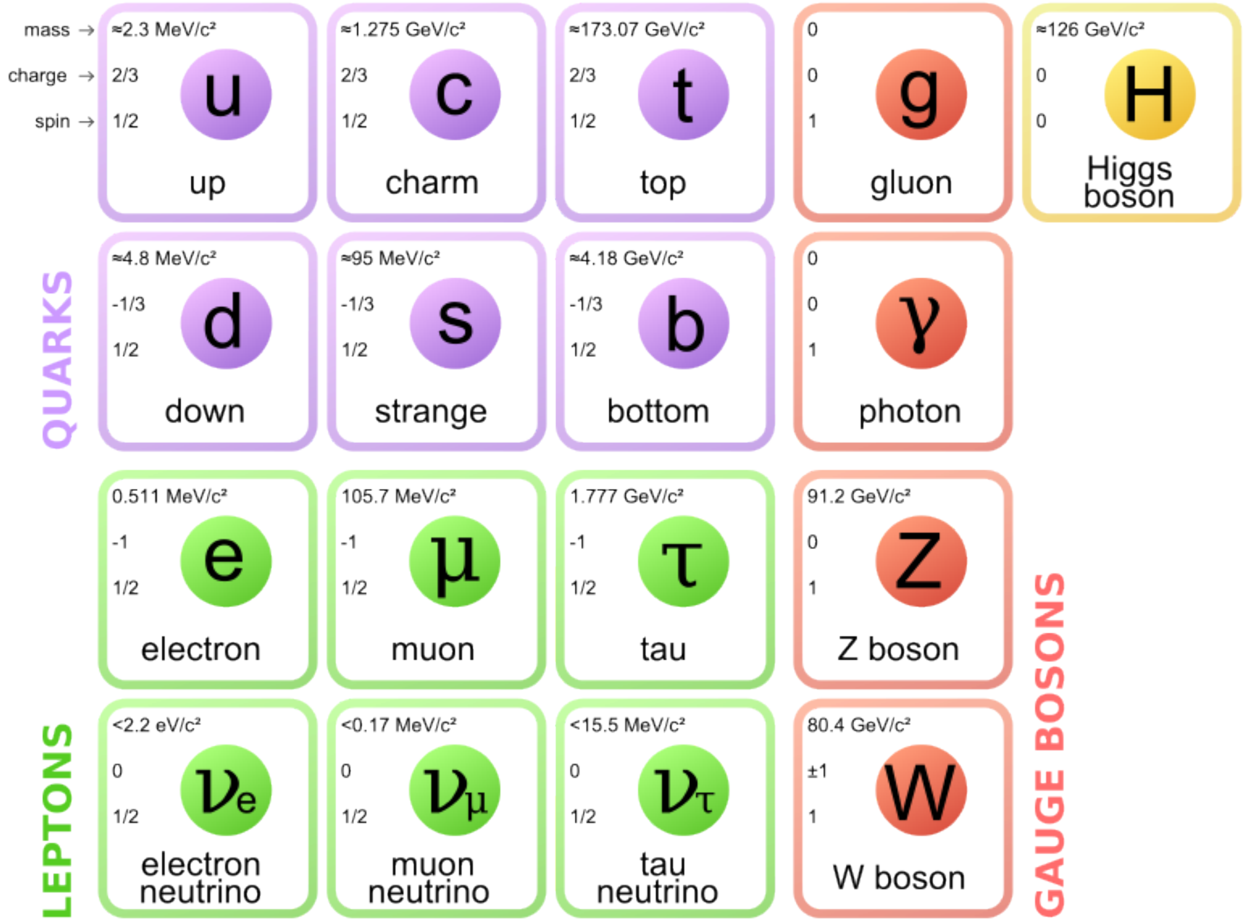
The Forces of Nature

- Gravitational interactions described by Einstein

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G_N}{c^4} T_{\mu\nu}$$

- Standard Model

electromagnetism
weak force
strong force
Higgs effect



mass → charge → spin →	≈2.3 MeV/c² 2/3 1/2 u up	≈1.275 GeV/c² 2/3 1/2 c charm	≈173.07 GeV/c² 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c² 0 0 H Higgs boson
QUARKS	≈4.8 MeV/c² -1/3 1/2 d down	≈95 MeV/c² -1/3 1/2 s strange	≈4.18 GeV/c² -1/3 1/2 b bottom	0 0 1 γ photon	
	0.511 MeV/c² -1 1/2 e electron	105.7 MeV/c² -1 1/2 μ muon	1.777 GeV/c² -1 1/2 τ tau	91.2 GeV/c² 0 1 Z Z boson	GAUGE BOSONS
LEPTONS	<2.2 eV/c² 0 1/2 ν_e electron neutrino	<0.17 MeV/c² 0 1/2 ν_μ muon neutrino	<15.5 MeV/c² 0 1/2 ν_τ tau neutrino	80.4 GeV/c² ±1 1 W W boson	

The Forces of Nature

- Gravitational interactions described by Einstein

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G_N}{c^4} T_{\mu\nu}$$

- Non-gravitational interactions are not encoded as geometry

Theorem [Coleman–Mandula]: symmetry group in 4 dimensions is Poincaré **x** internal

The Forces of Nature

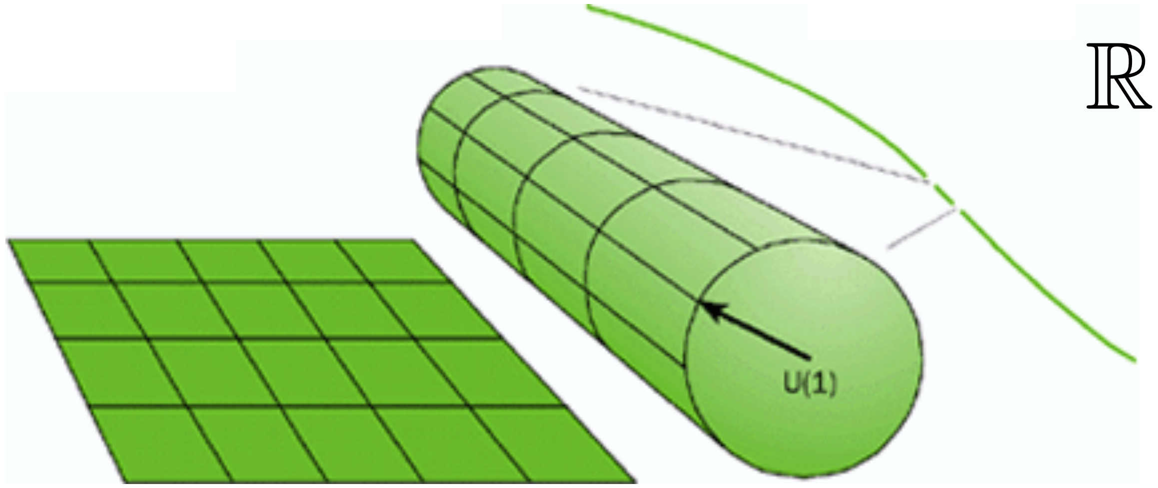
- Gravitational interactions described by Einstein

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G_N}{c^4} T_{\mu\nu}$$

- Non-gravitational interactions are not encoded as geometry

Theorem [Coleman–Mandula]: symmetry group in 4 dimensions is Poincaré **x** internal

- Clever loophole: internal symmetries may arise from higher dimensional geometry



$$\mathbb{R}^{1,3} \times S^1$$

Kaluza–Klein: 5d Einstein equations give 4d Einstein + Maxwell equations

Geometric Engineering

- Higher dimensional objects in string theory (branes) on which QFTs live

- Ten dimensional theory is consistent

- Ansatz for the geometry is $\mathcal{M}_{10} = \mathbb{R}^{1,3} \times \mathrm{CY}_3$

- Properties of Calabi–Yau determine physics in four dimensions

Example: $N_g = \frac{1}{2}|\chi|$ in simplest heterotic compactification models

Candelas, Horowitz, Strominger, Witten (1985)
Greene, Kirklin, Miron, Ross (1986)

Geometric Engineering

- Higher dimensional objects in string theory (branes) on which QFTs live

- Ten dimensional theory is consistent

- Ansatz for the geometry is $\mathcal{M}_{10} = \underset{\mathrm{dS}_4}{\xcancel{\mathbb{R}^{1,3}}} \times \mathrm{CY}_3$

- Properties of Calabi–Yau determine physics in four dimensions

Example: $N_g = \frac{1}{2}|\chi|$ in simplest heterotic compactification models

Candelas, Horowitz, Strominger, Witten (1985)
Greene, Kirklin, Miron, Ross (1986)

The Real World

- String theory supplies a framework for quantum gravity
- We are beginning to understand black holes and holography
- String theory is also an organizing principle for mathematics
- Finding our universe among the myriad of possible consistent realizations of a four dimensional low-energy limit of string theory is the **vacuum selection problem**
- Most vacua are *false* in that they do not resemble Nature at all
- Among the landscape of possibilities, we do not have even one solution that reproduces all the particle physics and cosmology we know

The Unreal World

- The objective is to obtain the real world from a string compactification

- We would happily settle for a modestly unreal world

$\mathcal{N}=1$ supersymmetry in 4 dimensions
$G = SU(3)_C \times SU(2)_L \times U(1)_Y$
Matter in chiral representations of G : $(\mathbf{3},\mathbf{2})_{\frac{1}{6}},(\overline{\mathbf{3}},\mathbf{1})_{-\frac{2}{3}},(\overline{\mathbf{3}},\mathbf{1})_{\frac{1}{3}},(\mathbf{1},\mathbf{2})_{\pm\frac{1}{2}},(\mathbf{1},\mathbf{1})_1,(\mathbf{1},\mathbf{1})_0$
Superpotential $W \supset \lambda^{ij}\phi\overline{\psi}^i_L\psi^j_R$
Three copies of matter such that λ^{ij} not identical
Consistent with cosmology

The Unreal World

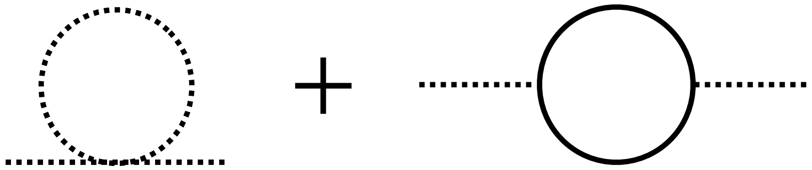
- The objective is to obtain the real world from a string compactification
- We would happily settle for a modestly unreal world

$\mathcal{N} = 1$ supersymmetry in 4 dimensions

No experimental evidence so far!

$$Q|\lambda\rangle \sim |\lambda \pm \frac{1}{2}\rangle$$

$$|\text{boson}\rangle \longleftrightarrow |\text{fermion}\rangle$$



$$m_H \ll m_P$$

The Unreal World

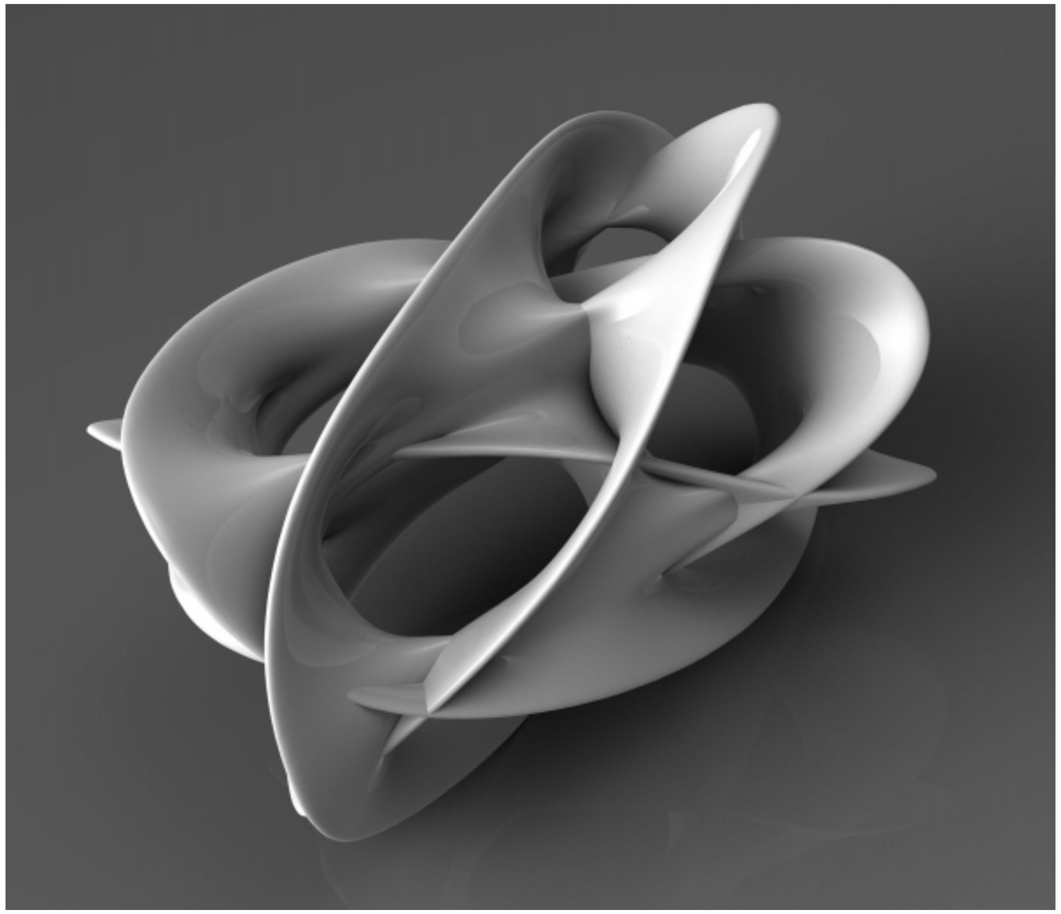
- The objective is to obtain the real world from a string compactification

- We would happily settle for a modestly unreal world

$\mathcal{N} = 1$ supersymmetry in 4 dimensions

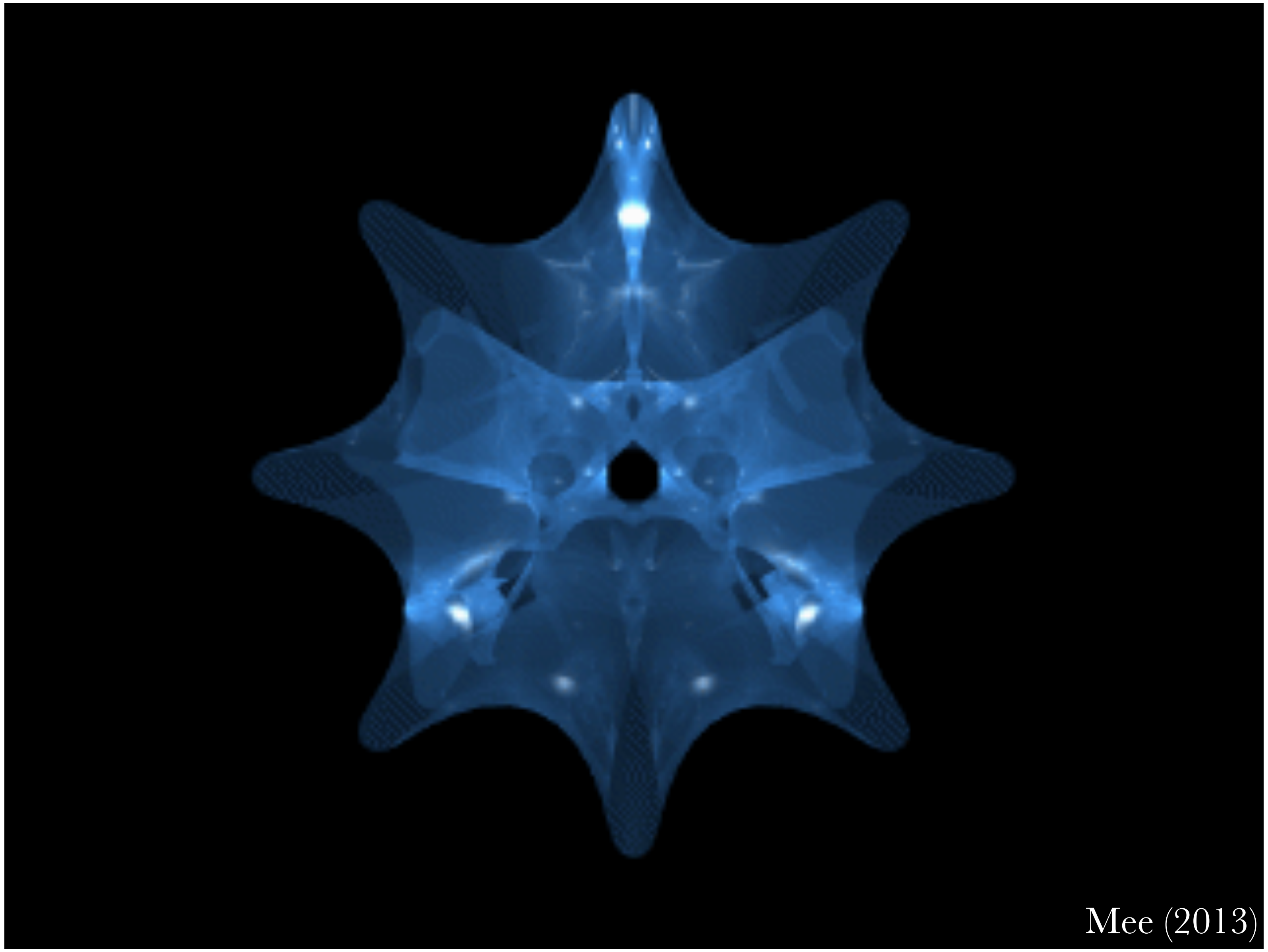
Because it is Ricci flat, the Calabi–Yau geometry ensures 4d supersymmetry

Use topological and geometric features of the Calabi–Yau to recover aspects of the real world



PREDICTING A CALABI~YAU'S TOPOLOGICAL INVARIANTS

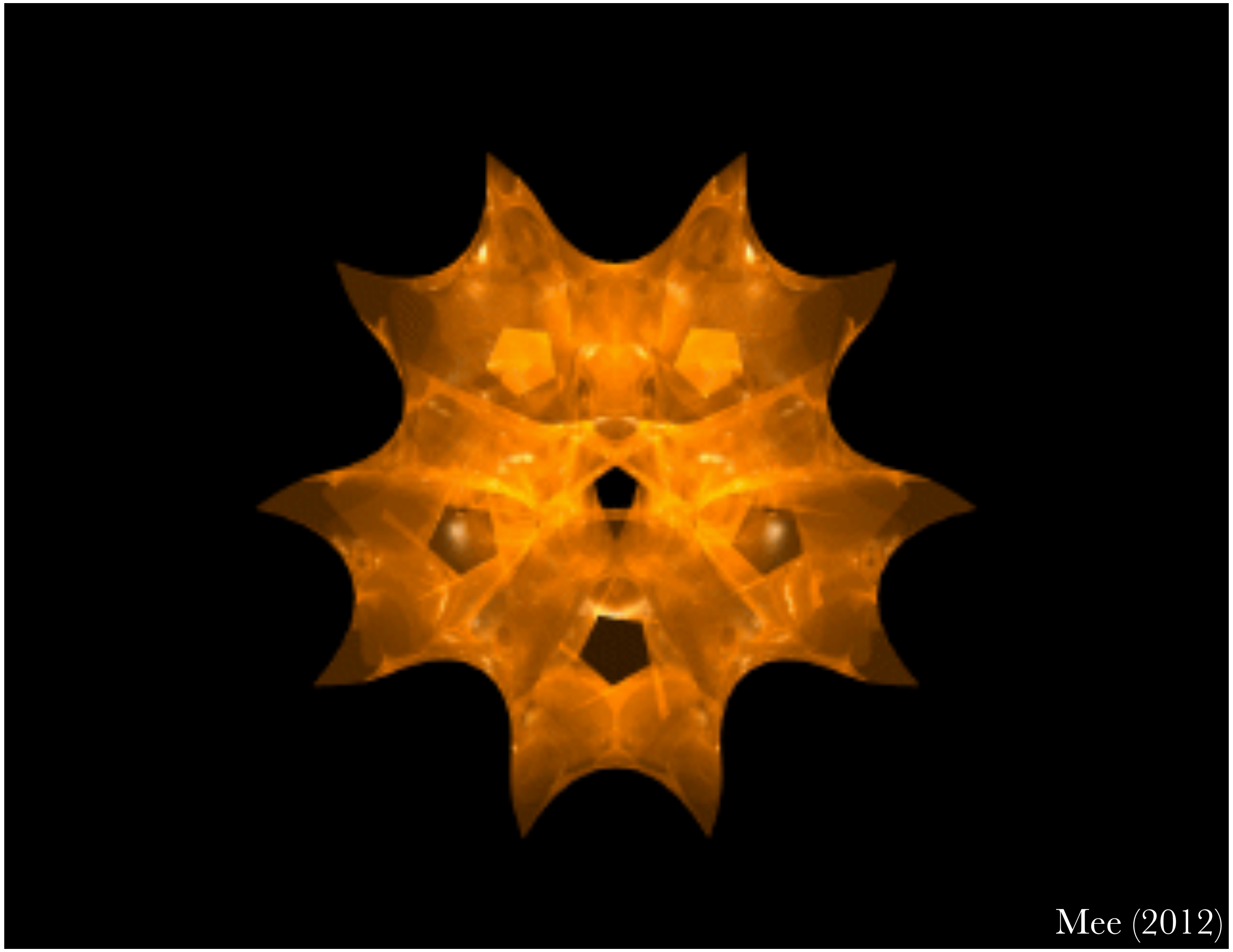
Calabi–Yau



Mee (2013)

$w^4 + x^4 + y^4 + z^4 = 0 \subset \mathbb{P}^3$

Calabi–Yau



Mee (2012)

$u^5 + v^5 + x^5 + y^5 + z^5 = 0 \subset \mathbb{P}^4$

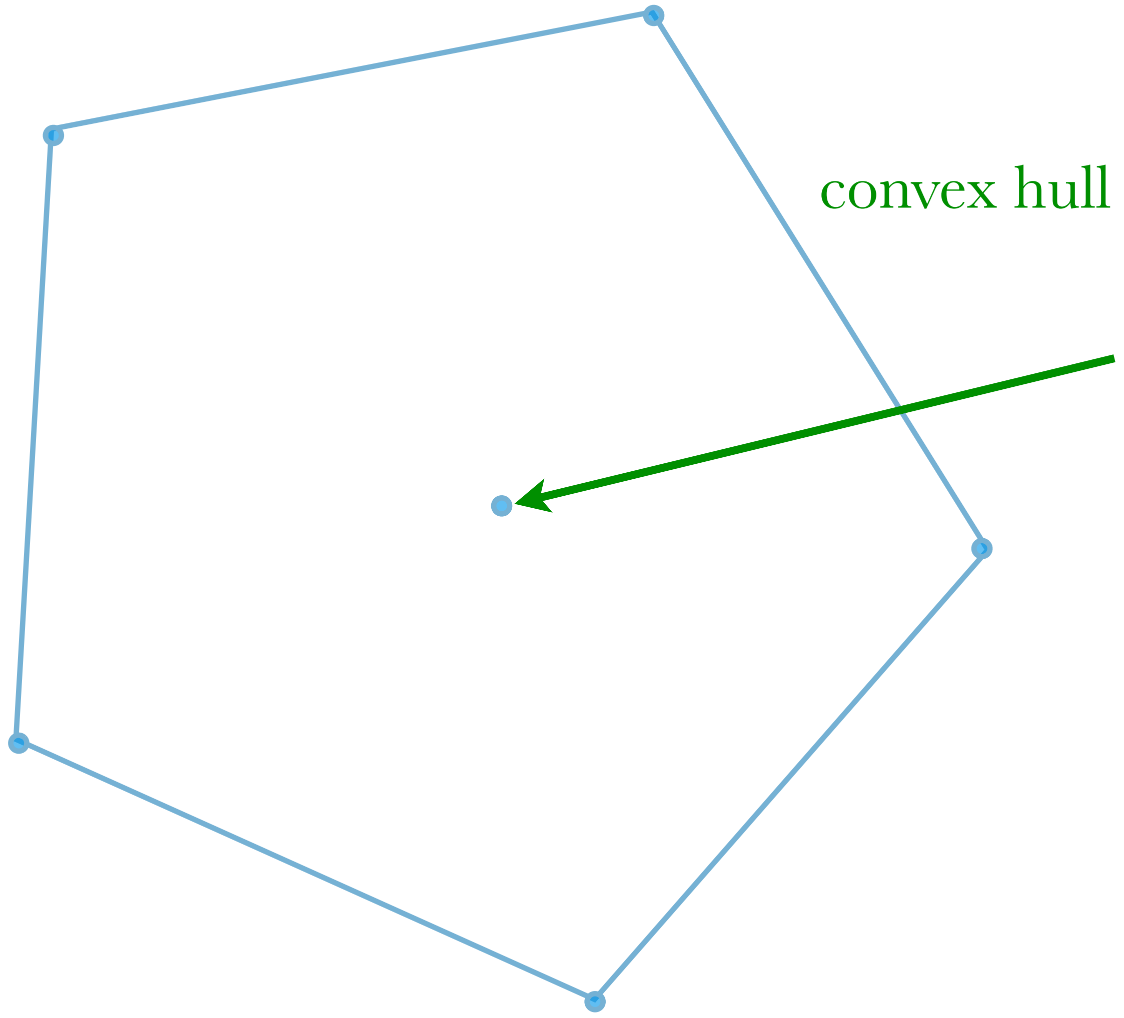
There is a nowhere vanishing holomorphic n-form

The canonical bundle is trivial

There is a Kähler metric with holonomy in $SU(n)$

Reflexive Polytopes

- Starting from a reflexive polytope, one can build a toric Calabi–Yau via methods of Batyrev, Borisov

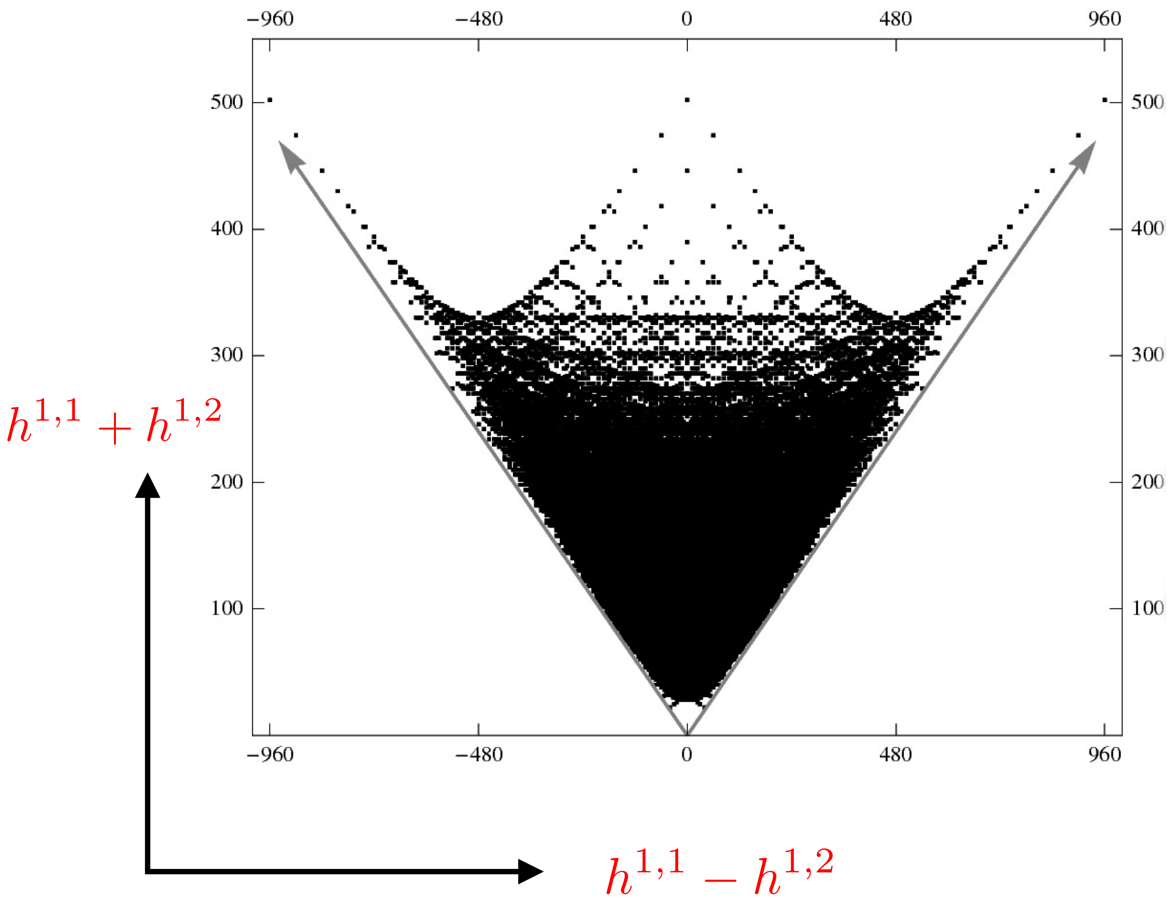


convex hull of finitely many lattice points

there is a single interior lattice point

Reflexive Polytopes Catalogued

- Starting from a reflexive polytope, one can build a toric Calabi–Yau via methods of Batyrev, Borisov
- Kreuzer–Skarke obtained 473,800,776 reflexive polytopes that yield toric Calabi–Yau threefolds with 30,108 unique pairs of Hodge numbers

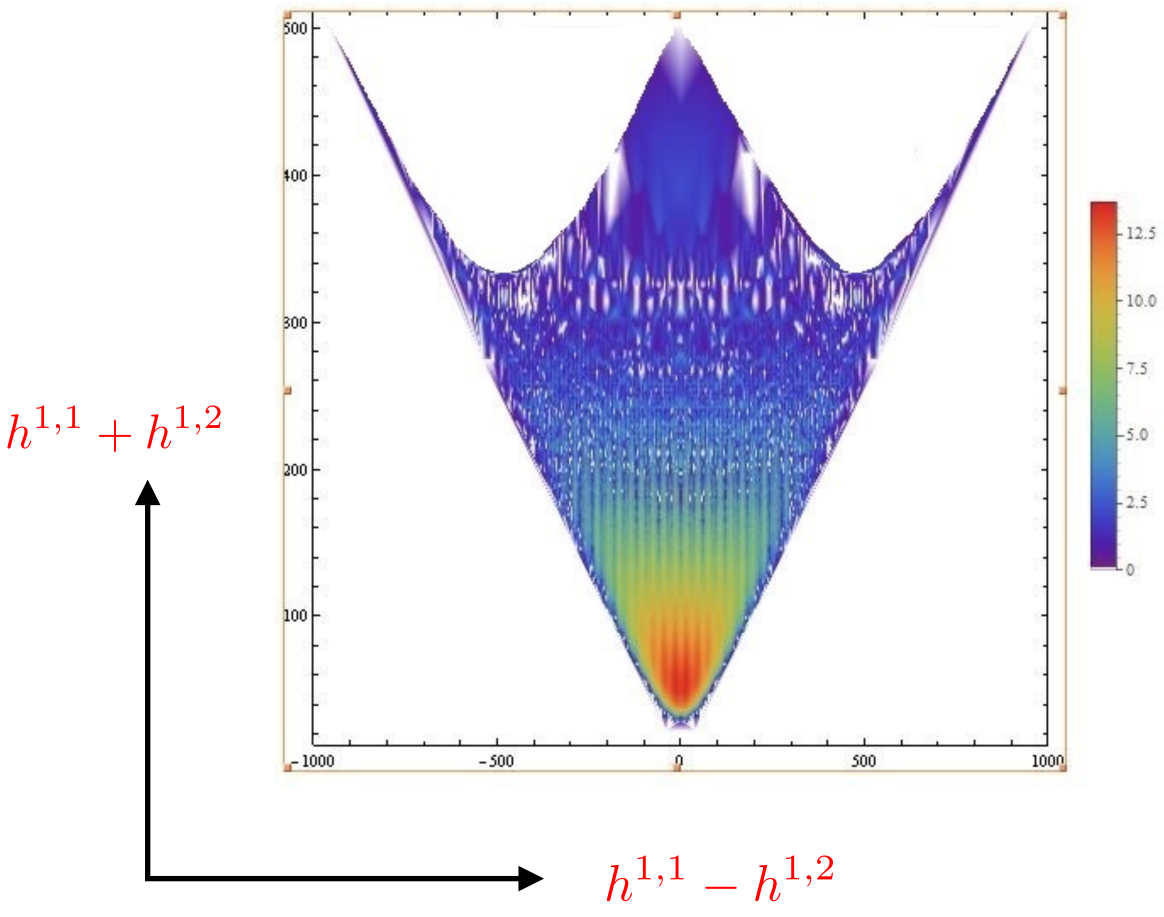


- Distribution of polytopes exhibits mirror symmetry

Reflexive Polytopes Catalogued

- Starting from a reflexive polytope, one can build a toric Calabi–Yau via methods of Batyrev, Borisov

- Kreuzer–Skarke obtained 473,800,776 reflexive polytopes that yield toric Calabi–Yau threefolds with 30,108 unique pairs of Hodge numbers



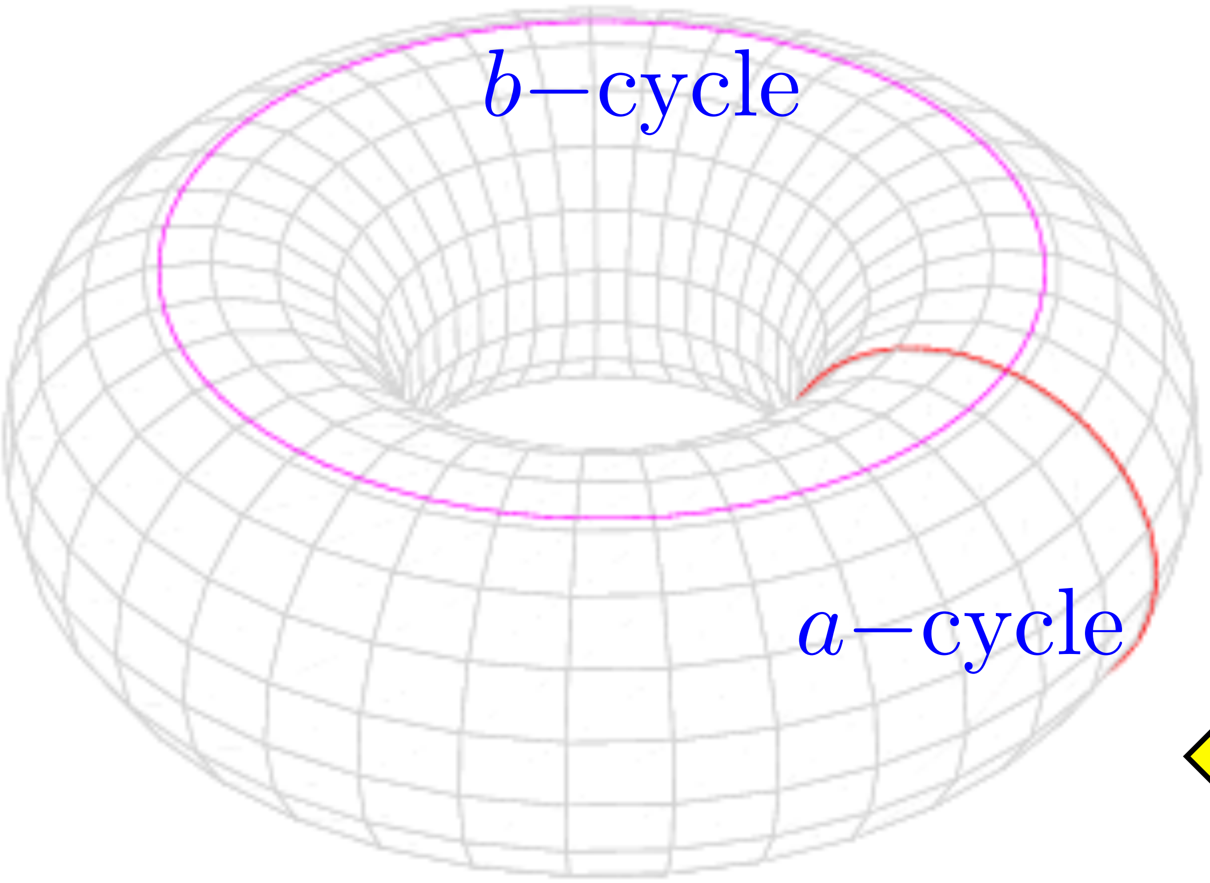
- Distribution of polytopes exhibits mirror symmetry

- The peak of the distribution is at $(h^{1,1}, h^{1,2}) = (27, 27)$

  There are 910,113 such polytopes

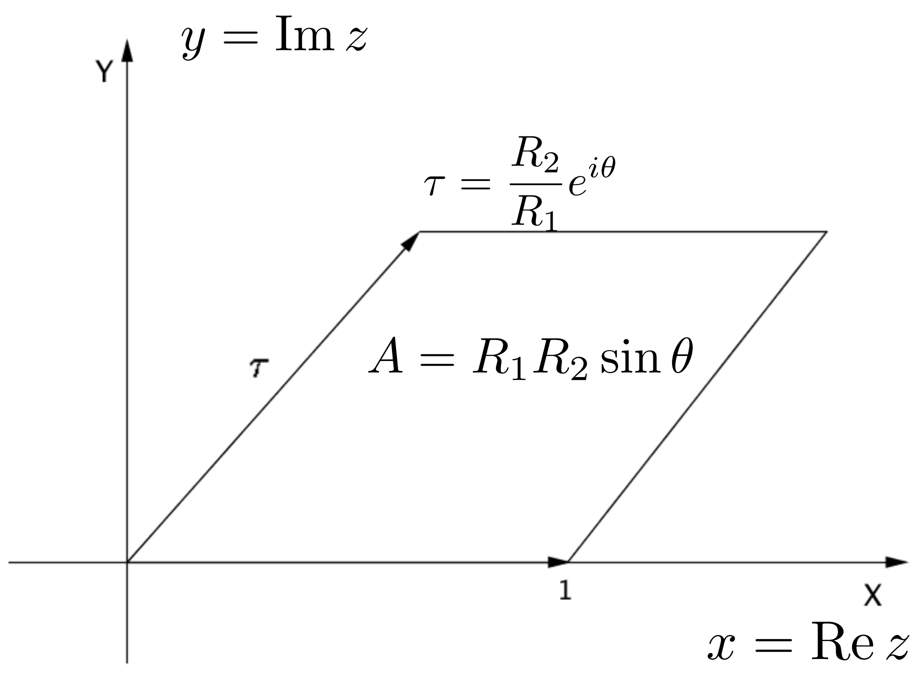
- Are there patterns in how the topological invariants are distributed?

Torus



Flat, but has non-trivial homotopy

There are non-contractible cycles



$$\mathrm{d}s^2 = R_1^2 dx^2 + R_2^2 dy^2 + 2R_1 R_2 \cos\theta dx\, dy$$

Kähler parameter: area A **size**

complex structure parameter: τ **shape**

Moduli of CY$_3$

- Geometrical moduli enumerated by number of embedded two-spheres and three-spheres

$$\begin{array}{cccccccc}
 & & & 1 & & & & b_0\\
 & & 0 & & 0 & & & b_1\\
 & 0 & & h^{1,1} & & 0 & & b_2\\
1 & & h^{1,2} & & h^{2,1} & & 1 & b_3\\
 & 0 & & h^{2,2} & & 0 & & b_4\\
 & & 0 & & 0 & & & b_5\\
 & & & 1 & & & & b_6
\end{array}$$

$$h^{p,q} = \dim\ H^{p,q}$$

$$b_k = \dim\ H^k = \sum_{p+q=k} h^{p,q}$$

$$h^{p,q} = h^{q,p} \qquad \text{(complex conjugation)}$$

$$h^{p,q} = h^{n-p,n-q} \qquad \text{(Poincaré duality)}$$

$$\chi = \sum_{p,q} (-1)^{p+q} h^{p,q}$$

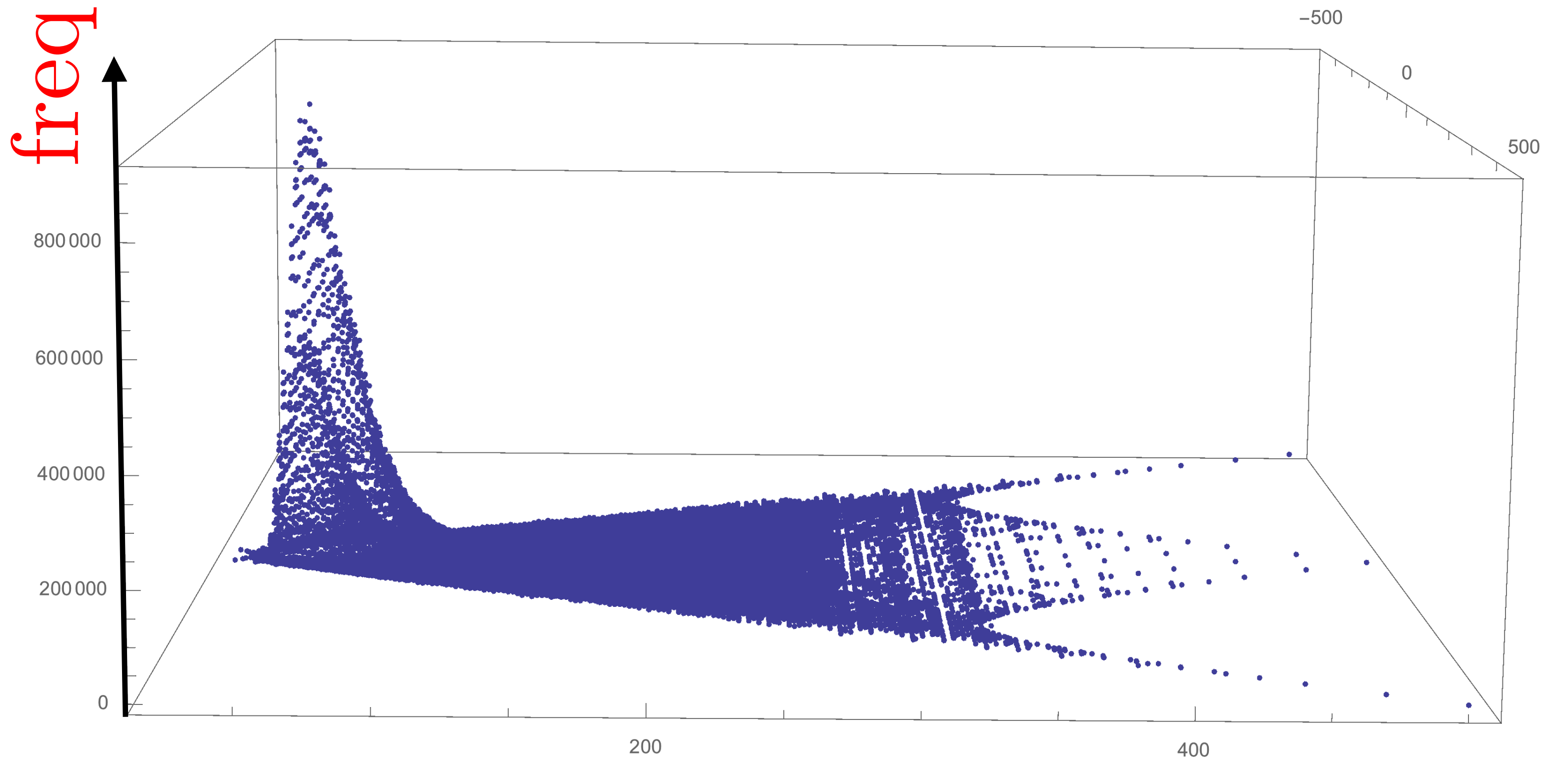
$h^{1,2} = \frac{b_3}{2} - 1$ **complex structure moduli**, counts the number of three-cycles

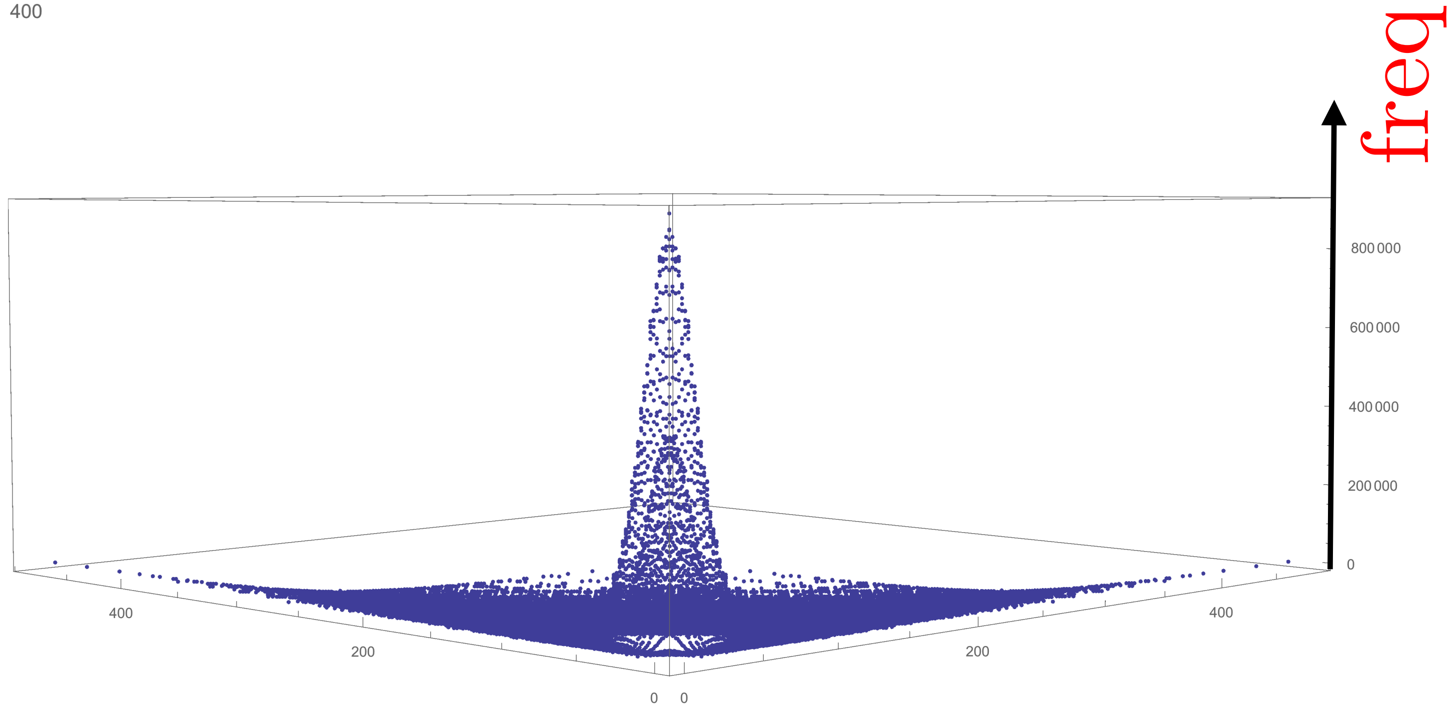
$h^{1,1} = b_2$ **Kähler moduli**, counts the number of two-cycles and four-cycles

$\chi = 2(h^{1,1} - h^{1,2})$ **Euler characteristic**, $N_g = \frac{1}{2}|\chi|$

- **Mirror symmetry** says that we can rotate the Hodge diamond by π/2 and get a new Calabi–Yau with $h^{1,1} \leftrightarrow h^{1,2}$

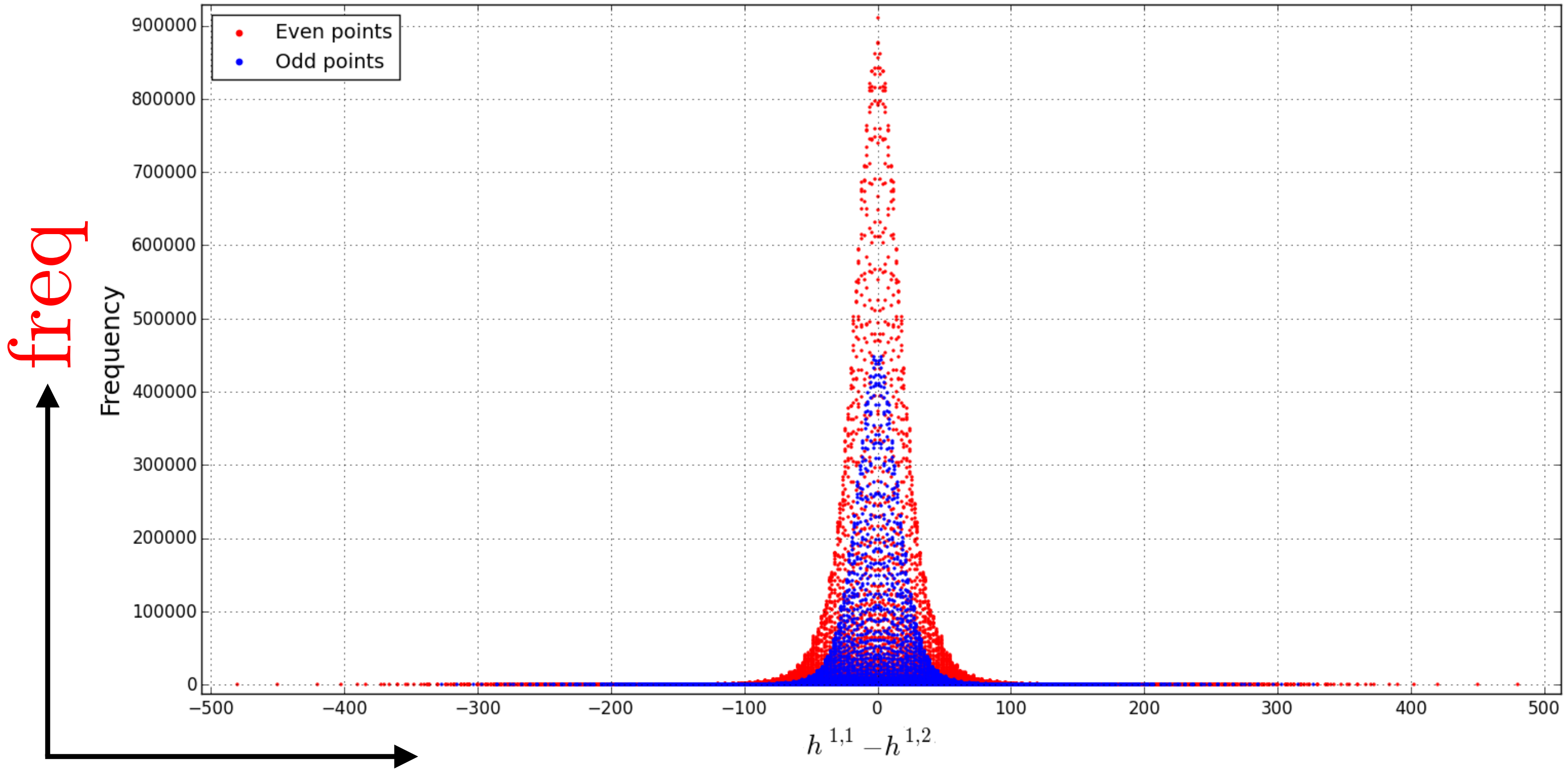
3d Plots of Polytope Data



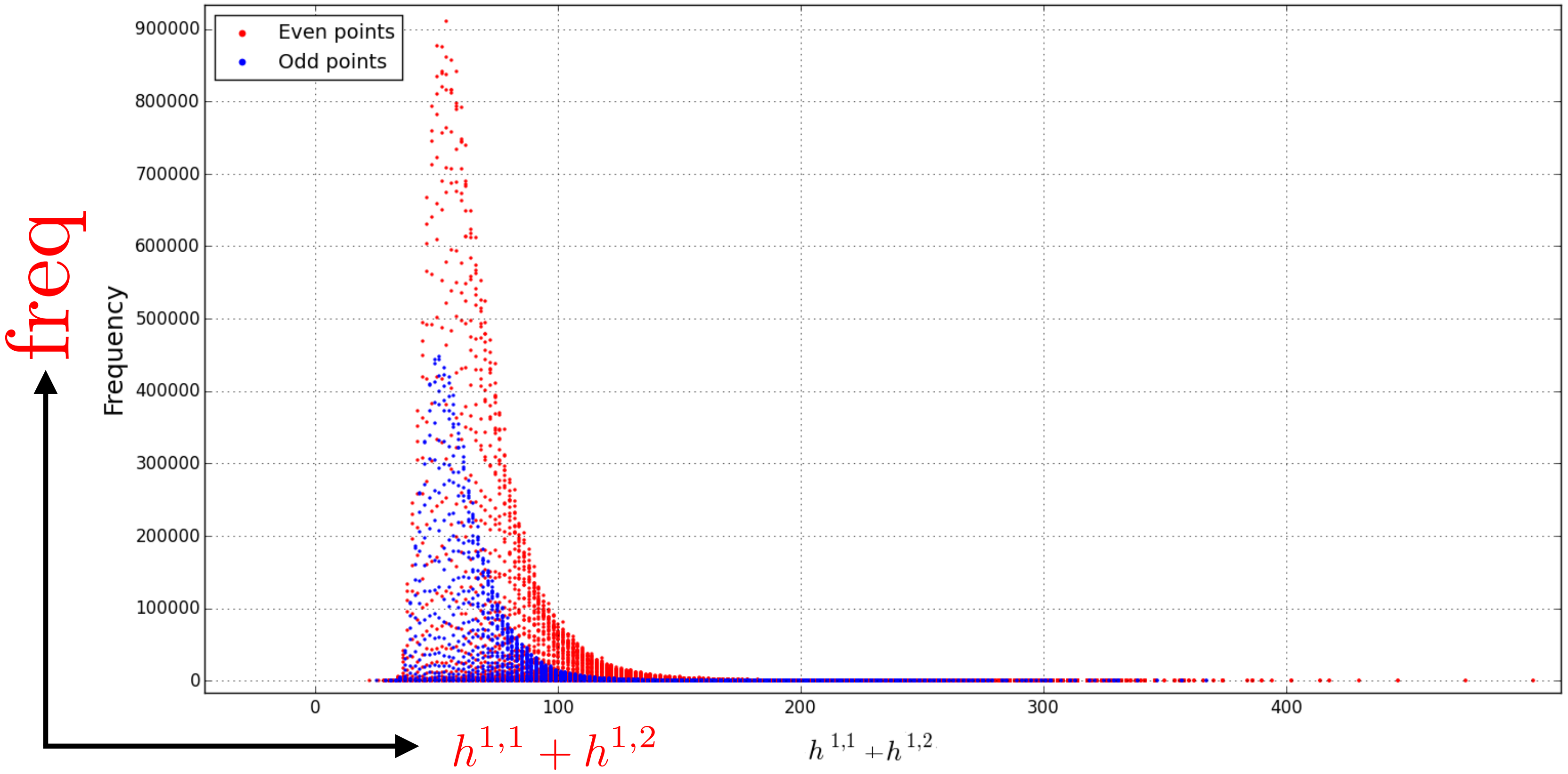


He, VJ, Pontiggia (2015)

Patterns in CY Distributions



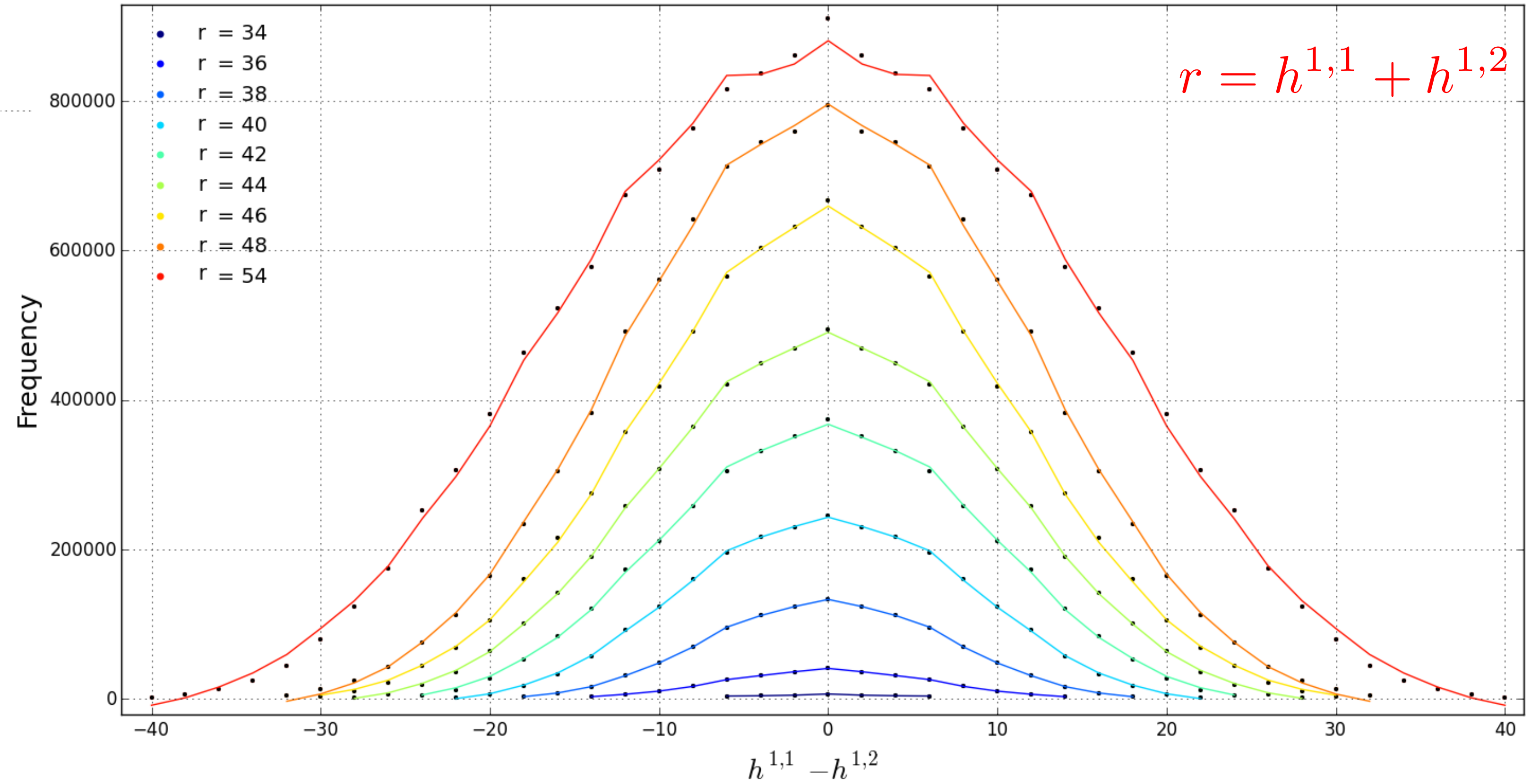
$h^{1,1} - h^{1,2}$



$h^{1,1} + h^{1,2}$

He, VJ, Pontiggia (2015)

Patterns in CY Distributions



$r = h^{1,1} + h^{1,2}$

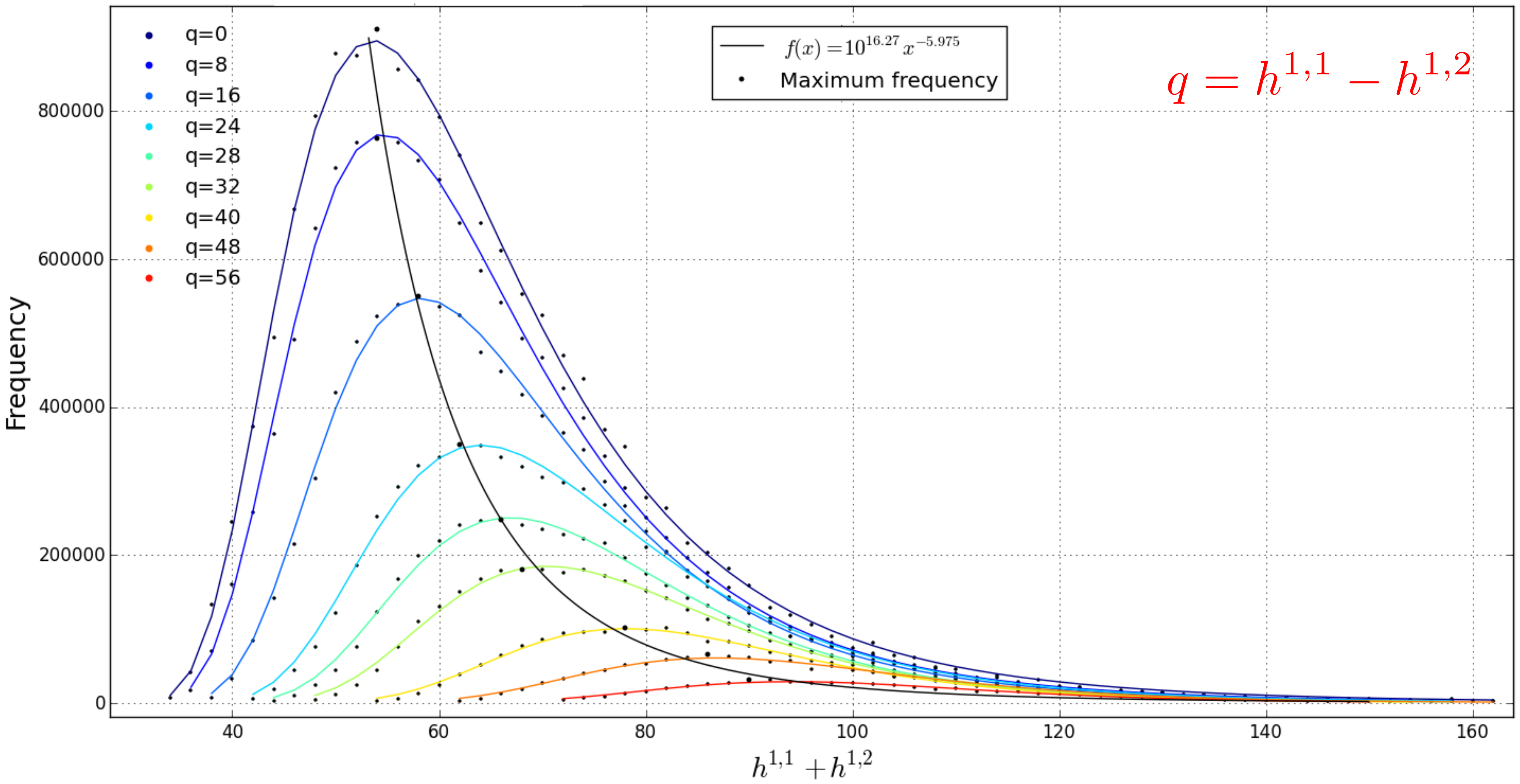
Pseudo-Voigt distribution

sum of Gaussian and Cauchy

$$(1-\alpha)\frac{A}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}+\alpha\frac{A}{\pi}\left[\frac{\sigma^2}{(x-\mu)^2+\sigma^2}\right]$$

Planck distribution

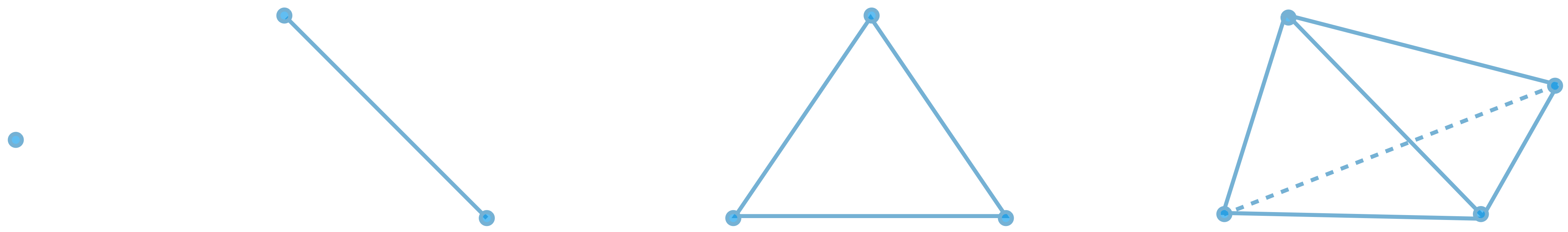
$$\frac{A}{x^n}\frac{1}{e^{b/(x-c)}-1}$$



$q = h^{1,1} - h^{1,2}$

He, VJ, Pontiggia (2015)

From Polytopes to Geometries



- A **triangulation** of $\mathcal{P}$ is a partition into simplices such that:

  the union of all simplices is $\mathcal{P}$

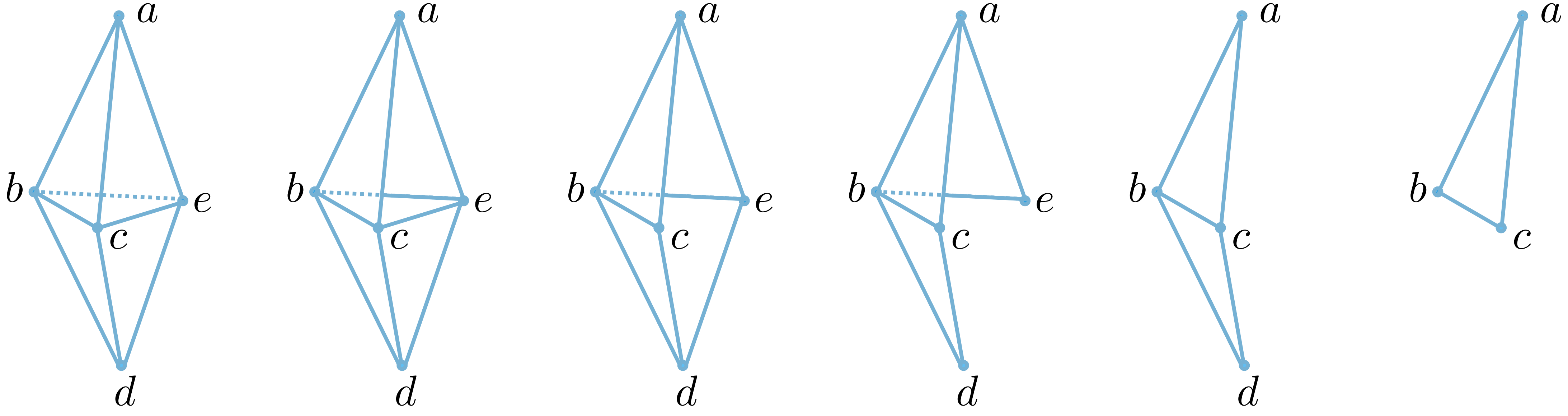
  the intersection of any pair is a (possibly empty) common face

- From triangulation, we construct the Stanley–Reisner ring

- Unique rings correspond to different Calabi–Yau geometries

- For each, we have topological data, intersection form, Kähler cone

Example: S^2



$I_\Delta = (ad, bce)$ minimal non-faces

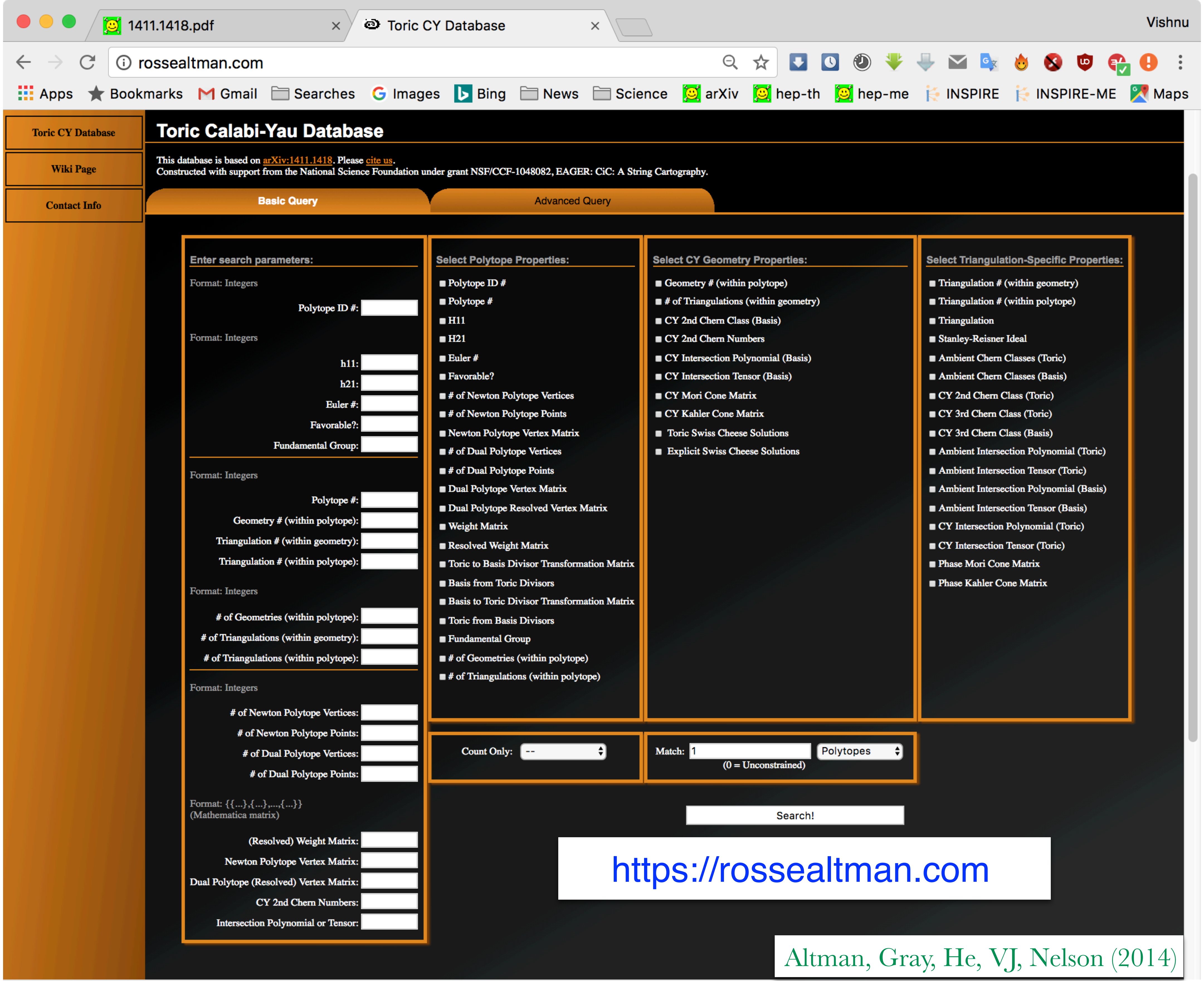
$\mathbb{K}_\Delta = \mathbb{K}[a, b, c, d, e]/I_\Delta$ Stanley–Reisner ring

Homeomorphic to two-sphere

From Polytopes to Geometries

- Every triangulation of a reflexive polytope can yield a Calabi–Yau
- We do not know how many toric Calabi–Yau geometries there are
- Different triangulations of the same polytope are expected, in general, to give different Calabi–Yau manifolds
- In principle, triangulations of different polytopes can give the same Calabi–Yau manifold
- The Calabi–Yau inherits topological invariants from the polytope
- 16 polytopes in $\mathbb{R}^2$ give rise to elliptic curves (Calabi–Yau onefolds)
  4319 polytopes in $\mathbb{R}^3$ give rise to K3 (Calabi–Yau twofolds)
  473800776 polytopes in $\mathbb{R}^4$ give rise to at least 30108 Calabi–Yau threefolds

A Calabi–Yau Database



Toric Calabi-Yau Database

This database is based on arXiv:1411.1418. Please cite us.
Constructed with support from the National Science Foundation under grant NSF/CCF-1048082, EAGER: CiC: A String Cartography.

Enter search parameters:

Format: Integers

Polytope ID #:

Format: Integers

h11:
h21:
Euler #:
Favorable?:
Fundamental Group:

Format: Integers

Polytope #:
Geometry # (within polytope):
Triangulation # (within geometry):
Triangulation # (within polytope):

Format: Integers

# of Geometries (within polytope):
# of Triangulations (within geometry):
# of Triangulations (within polytope):

Format: Integers

# of Newton Polytope Vertices:
# of Newton Polytope Points:
# of Dual Polytope Vertices:
# of Dual Polytope Points:

Format: {{...},{...},...,{...}}
(Mathematica matrix)

(Resolved) Weight Matrix:
Newton Polytope Vertex Matrix:
Dual Polytope (Resolved) Vertex Matrix:
CY 2nd Chern Numbers:
Intersection Polynomial or Tensor:

Select Polytope Properties:

- ■ Polytope ID #
- ■ Polytope #
- ■ H11
- ■ H21
- ■ Euler #
- ■ Favorable?
- ■ # of Newton Polytope Vertices
- ■ # of Newton Polytope Points
- ■ Newton Polytope Vertex Matrix
- ■ # of Dual Polytope Vertices
- ■ # of Dual Polytope Points
- ■ Dual Polytope Vertex Matrix
- ■ Dual Polytope Resolved Vertex Matrix
- ■ Weight Matrix
- ■ Resolved Weight Matrix
- ■ Toric to Basis Divisor Transformation Matrix
- ■ Basis from Toric Divisors
- ■ Basis to Toric Divisor Transformation Matrix
- ■ Toric from Basis Divisors
- ■ Fundamental Group
- ■ # of Geometries (within polytope)
- ■ # of Triangulations (within polytope)

Select CY Geometry Properties:

- ■ Geometry # (within polytope)
- ■ # of Triangulations (within geometry)
- ■ CY 2nd Chern Class (Basis)
- ■ CY 2nd Chern Numbers
- ■ CY Intersection Polynomial (Basis)
- ■ CY Intersection Tensor (Basis)
- ■ CY Mori Cone Matrix
- ■ CY Kahler Cone Matrix
- ■ Toric Swiss Cheese Solutions
- ■ Explicit Swiss Cheese Solutions

Select Triangulation-Specific Properties:

- ■ Triangulation # (within geometry)
- ■ Triangulation # (within polytope)
- ■ Triangulation
- ■ Stanley-Reisner Ideal
- ■ Ambient Chern Classes (Toric)
- ■ Ambient Chern Classes (Basis)
- ■ CY 2nd Chern Class (Toric)
- ■ CY 3rd Chern Class (Toric)
- ■ CY 3rd Chern Class (Basis)
- ■ Ambient Intersection Polynomial (Toric)
- ■ Ambient Intersection Tensor (Toric)
- ■ Ambient Intersection Polynomial (Basis)
- ■ Ambient Intersection Tensor (Basis)
- ■ CY Intersection Polynomial (Toric)
- ■ CY Intersection Tensor (Toric)
- ■ Phase Mori Cone Matrix
- ■ Phase Kahler Cone Matrix

Count Only: --

Match: 1 Polytopes
(0 = Unconstrained)

https://rossealtman.com

Altman, Gray, He, VJ, Nelson (2014)

CICYs

- Zero locus of a set of homogeneous polynomials over combined set of coordinates of projective spaces

$$X = \begin{array}{c} \mathbb{P}^{n_1} \\ \vdots \\ \mathbb{P}^{n_\ell} \end{array} \left(\begin{array}{ccc} q_1^1 & \cdots & q_K^1 \\ \vdots & \ddots & \vdots \\ q_1^\ell & \cdots & q_K^\ell \end{array} \right)_\chi$$

configuration matrix

$$\sum_r n_r - K = 3$$ complete intersection threefold

$$\sum_a q_a^r = n_r + 1\,, \quad \forall\ r \in \{1, \ldots, \ell\}$$

$$c_1 = 0$$

- K equations of multi-degree $q_a^r \in \mathbb{Z}_{\geq 0}$ embedded in $\mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_\ell}$

- **Example:** quintic $\mathbb{P}^4(5)_{-200}$

$$4 - 1 = 3$$

$$5 = 4 + 1$$

- Other examples:

$$\mathbb{P}^5(3,3)_{-144}\ ,\ \mathbb{P}^5(4,2)_{-176}\ ,\ \mathbb{P}^6(3,2,2)_{-144}\ ,\ \mathbb{P}^7(2,2,2,2)_{-128}$$

CICYs

- Tian–Yau manifold: $\begin{array}{c} \mathbb{P}^3 \\ \mathbb{P}^3 \end{array} \left(\begin{array}{ccc} 3 & 0 & 1 \\ 0 & 3 & 1 \end{array} \right)_{-18}$ $\Longleftrightarrow$ $\begin{array}{rcl} a^{\alpha\beta\gamma} w_\alpha w_\beta w_\gamma & = & 0 \\ b^{\alpha\beta\gamma} z_\alpha z_\beta z_\gamma & = & 0 \\ c^{\alpha\beta} w_\alpha z_\beta & = & 0 \end{array}$

$$h^{1,1} = 14 \ , \quad h^{1,2} = 23$$

freely acting $\mathbb{Z}_3$ quotient gives manifold with $\chi = -6$

central to early string phenomenology

- Transpose is Schön's manifold, also Calabi–Yau

$$\begin{array}{c} \mathbb{P}^2 \\ \mathbb{P}^2 \\ \mathbb{P}^1 \end{array} \left(\begin{array}{cc} 3 & 0 \\ 0 & 3 \\ 1 & 1 \end{array} \right)_{\chi = 0}$$

$$h^{1,1} = h^{1,2} = 19$$

- Can compute χ from configuration matrix

$$\frac{1}{3} \cdot 5 \cdot (5 - 5^3) = -200$$

$$\frac{1}{3} \cdot (4 \times 2) \cdot (6 - 4^3 - 2^3) = -176$$

$$\frac{1}{3} \cdot (3 \times 3) \cdot (6 - 3^3 - 3^3) = -144$$

$$\vdots$$

CICYs

- We have: 7890 configuration matrices

  Candelas, He, Hübsch, Lutken, Lynker, Schimmrigk, Berglund (1986-1990)

  1×1 to 12×15 with $q_a^r \in [0, 5]$

  266 distinct Hodge pairs $0 \leq h^{1,1} \leq 19 \ , \ 0 \leq h^{1,2} \leq 101$

  70 distinct Euler characters $\chi \in [-200, 0]$

  195 have freely acting symmetries, 37 different finite groups

  from $\mathbb{Z}_2$ to $\mathbb{Z}_8 \rtimes H_8$

  Braun (2010)

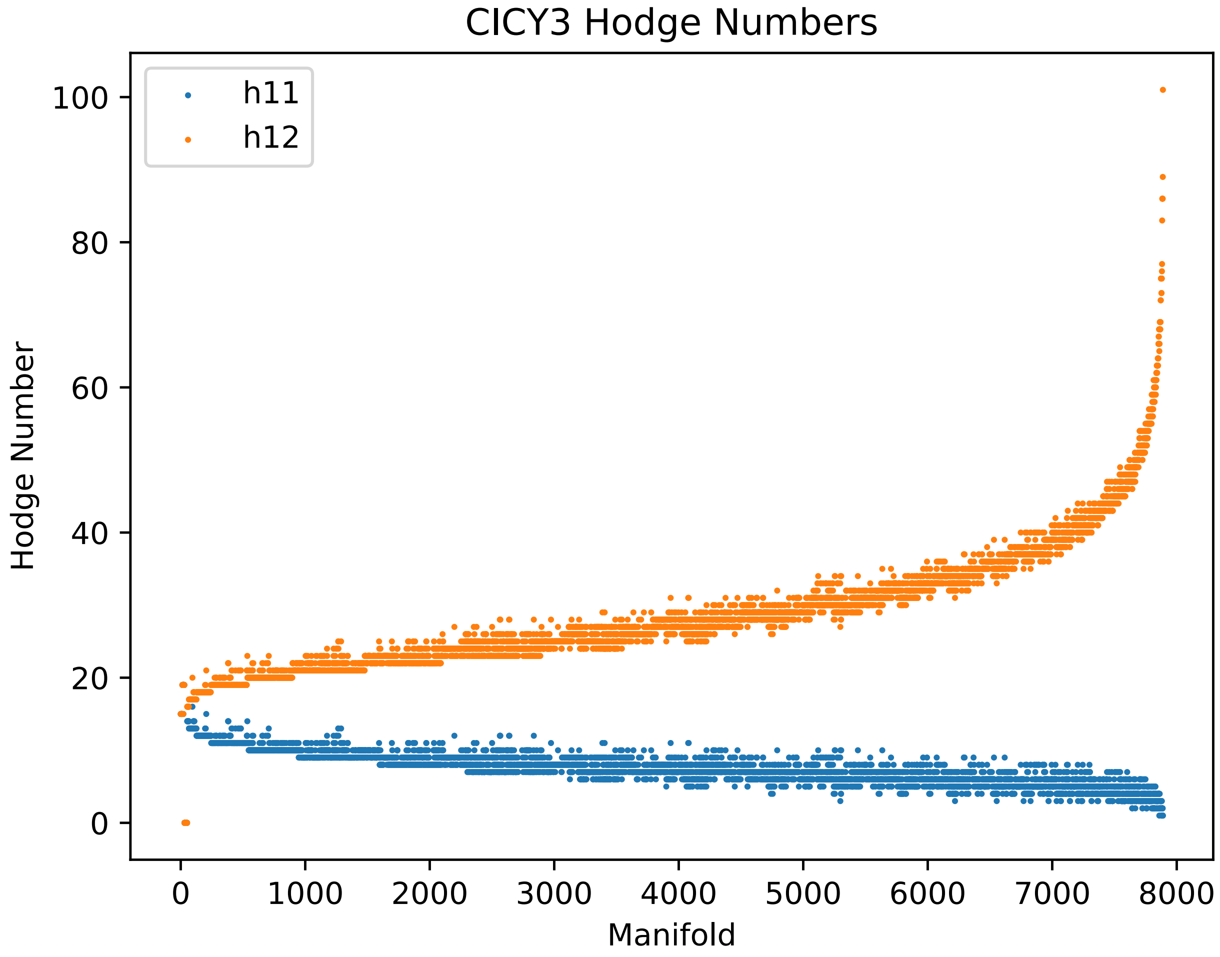
- By comparison, for fourfolds, there are 921497 CICYs

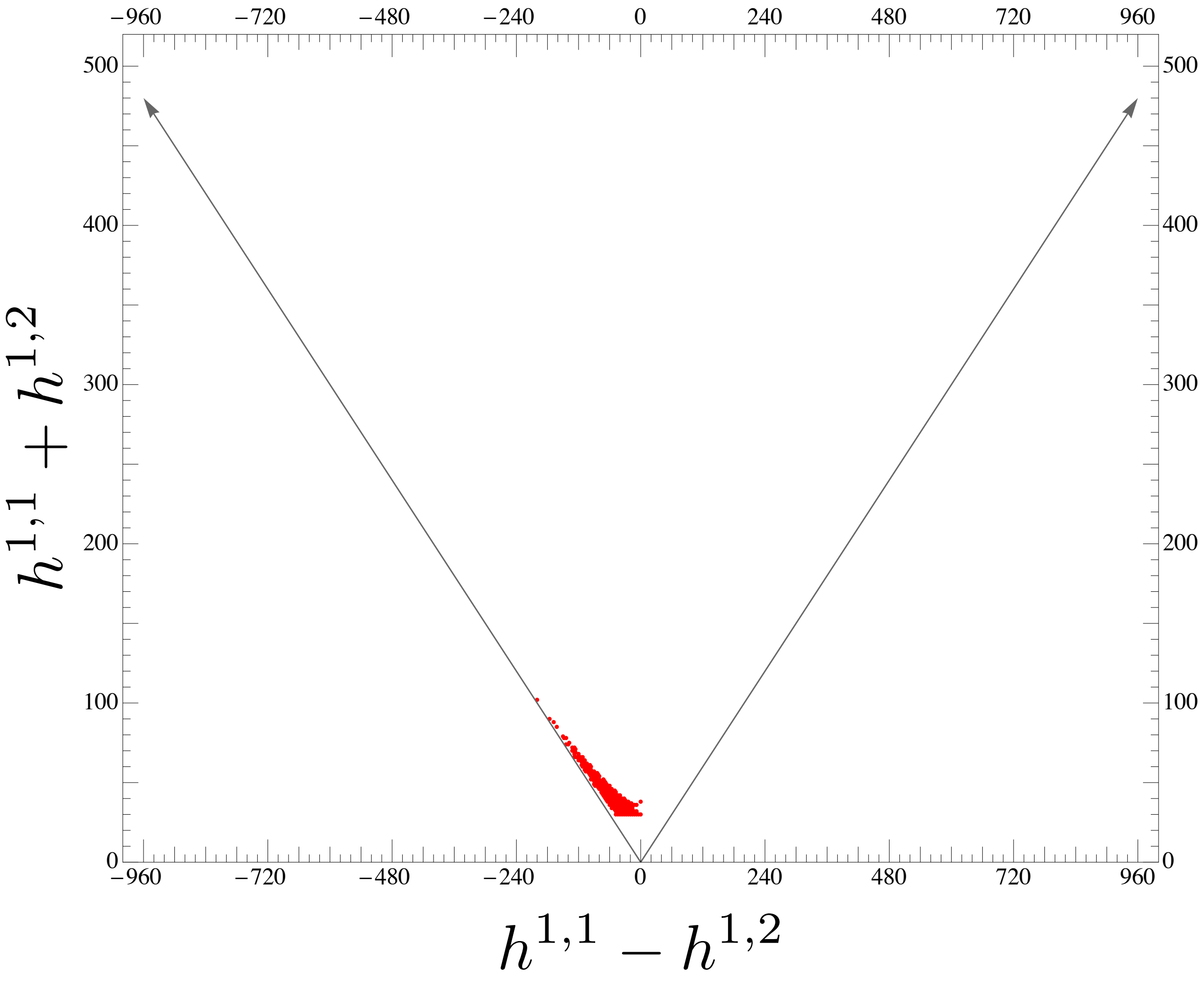
$$4h^{1,1} - 2h^{1,2} + 4h^{1,3} - h^{2,2} + 44 = 0$$

  Most of these are elliptically fibered

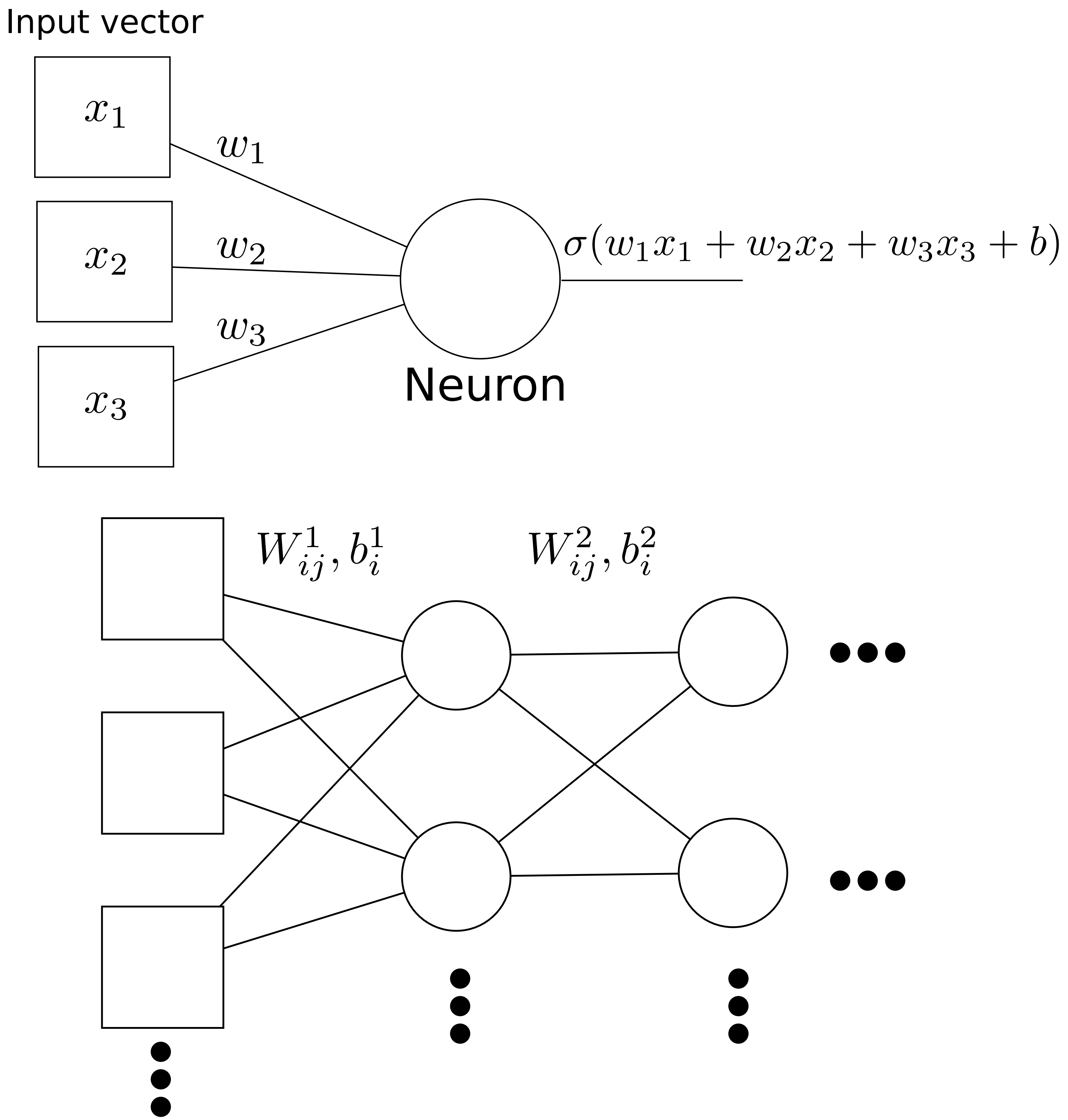
  Gray, Haupt, Lukas (2013)

CICY Hodge Numbers



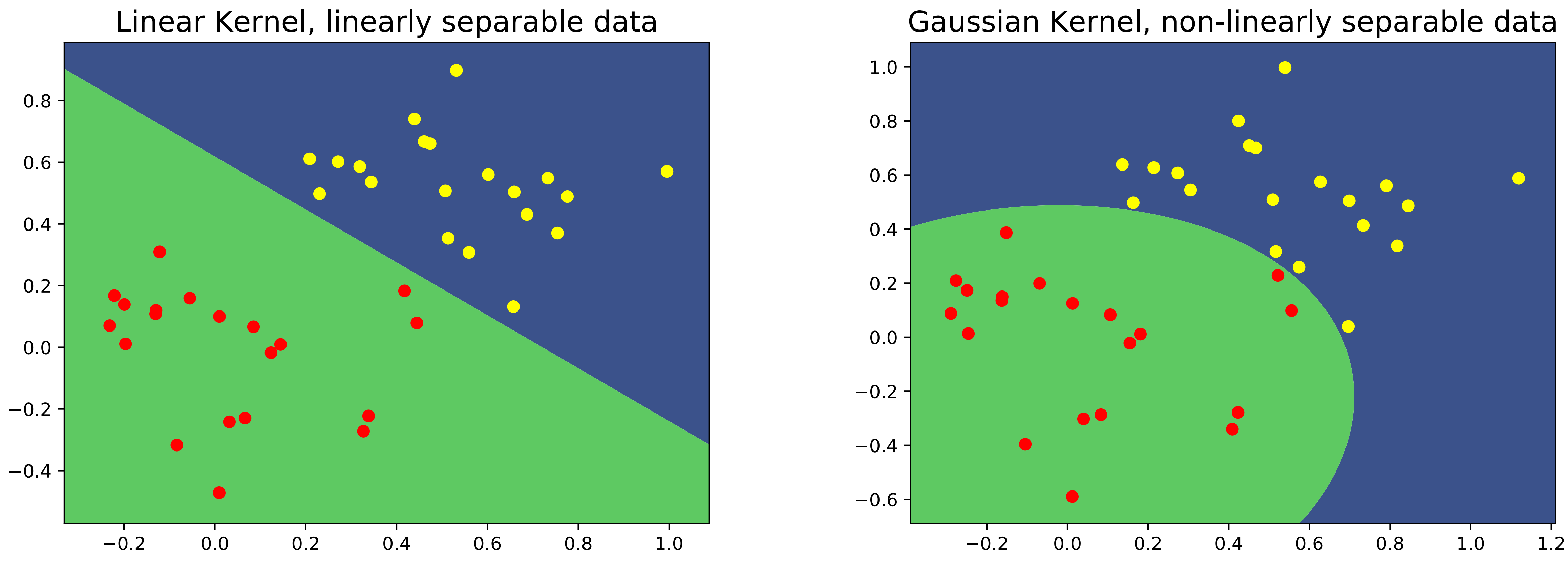


Feedforward Neural Networks



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.

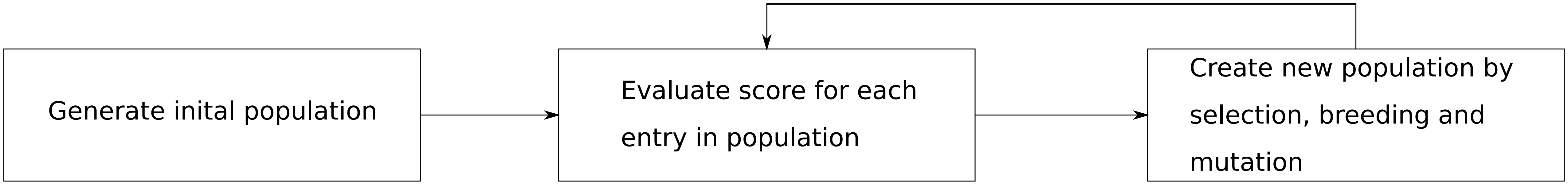
Support Vector Machines


Linear Kernel, linearly separable data

Gaussian Kernel, non-linearly separable data

SVM separation boundary calculated using our cvxopt implementation with a randomly generated data set.

Genetic Algorithms



Generate inital population

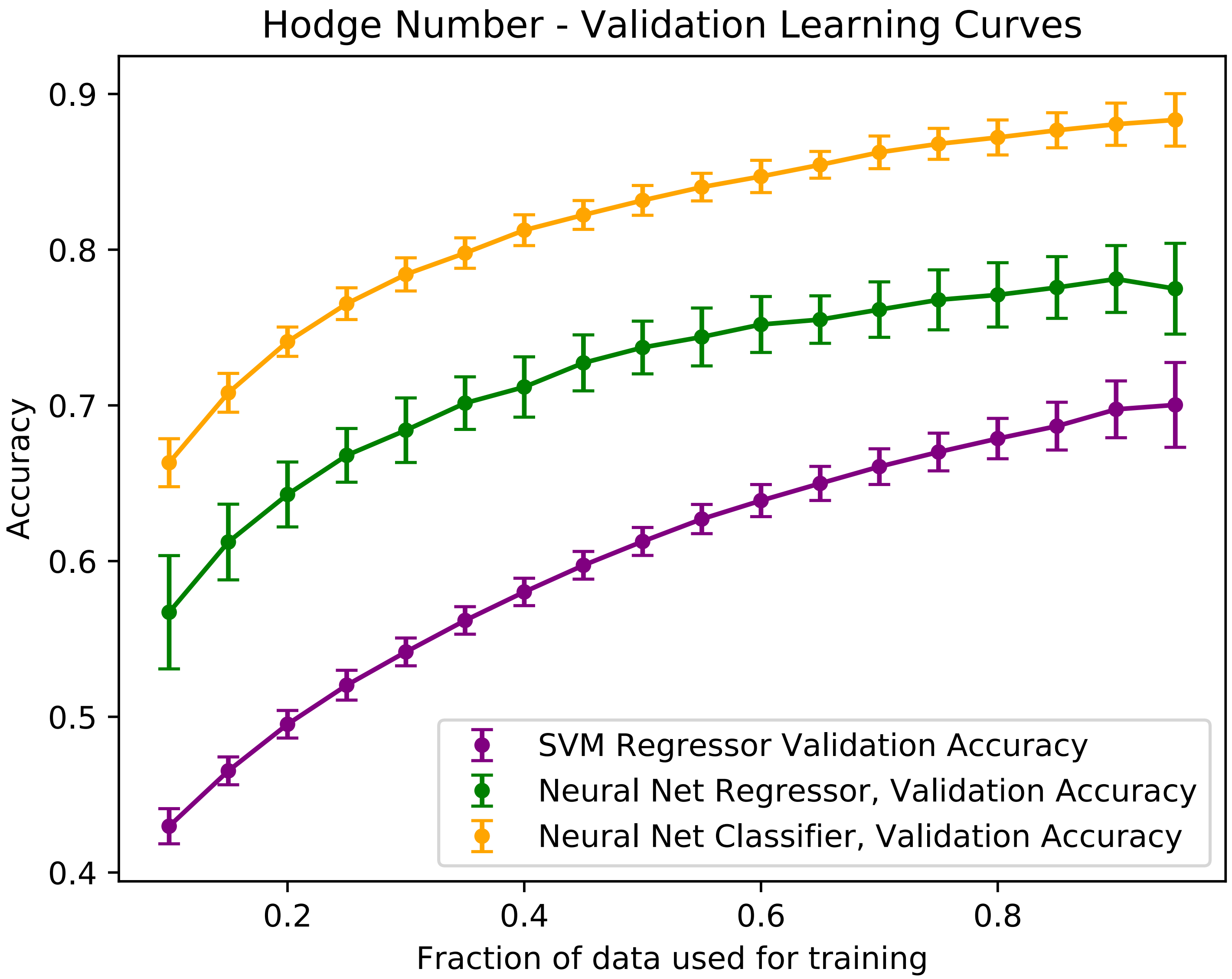
Evaluate score for each entry in population

Create new population by selection, breeding and mutation

Used to fix hyperparameters (*e.g.*, number of hidden layers and neurons in them, activation functions, learning rates and dropout) in neural network.

Machine Learning $h^{1,1}$

- Since we know $\chi = 2(h^{1,1} - h^{1,2})$ from intersection matrix, we choose to machine learn $h^{1,1} \in [0, 19]$
- Previous efforts discriminated large and small $h^{1,1}$
- Use Neural Network classifier/regressor and SVM regressor



Bull, He, VJ, Mishra (2018)

Machine Learning $h^{1,1}$

	Accuracy	RMS	R^2	WLB	WUB
SVM Reg	0.70 ± 0.02	**0.53 ± 0.06**	**0.78 ± 0.08**	0.642	0.697
NN Reg	0.78 ± 0.02	0.46 ± 0.05	0.72 ± 0.06	0.742	0.791
NN Class	**0.88 ± 0.02**	-	-	**0.847**	**0.886**

$$\text{RMS} := \left(\frac{1}{N} \sum_{i=1}^{N} (y_i^{pred} - y_i)^2 \right)^{1/2} \qquad R^2 := 1 - \frac{\sum_i (y_i - y_i^{pred})^2}{\sum_i (y_i - \bar{y})^2}$$

$$\omega_{\pm} := \frac{p + \frac{z^2}{2n}}{1 + \frac{z^2}{n}} \pm \frac{z}{1 + \frac{z^2}{n}} \left(\frac{p(1-p)}{n} + \frac{z^2}{4n^2} \right)^{1/2}$$

Wilson upper/lower bounds
(WUB/WLB)

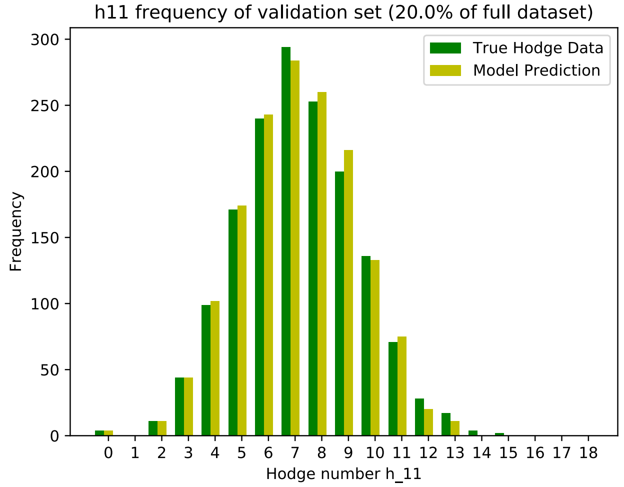
y_i	actual value
$\bar{y}$	average value
y_i^{pred}	predicted value
p	probability of successful prediction
z	probit
n	number of samples

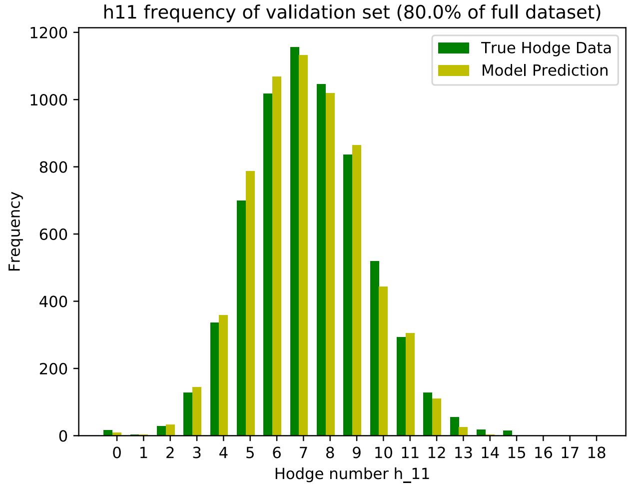
Bull, He, VJ, Mishra (2018)

Machine Learning $h^{1,1}$

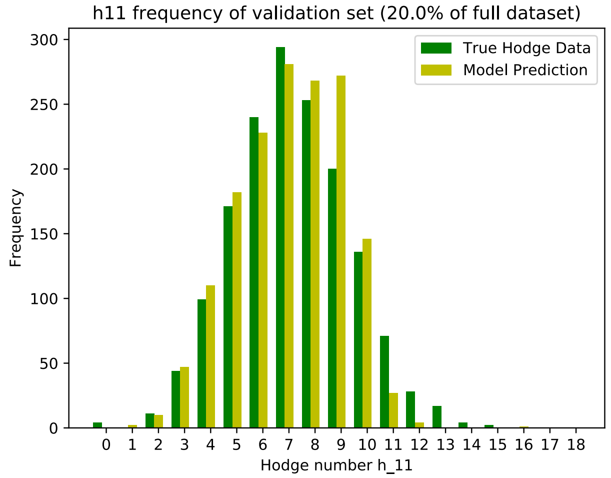
20% **80%**

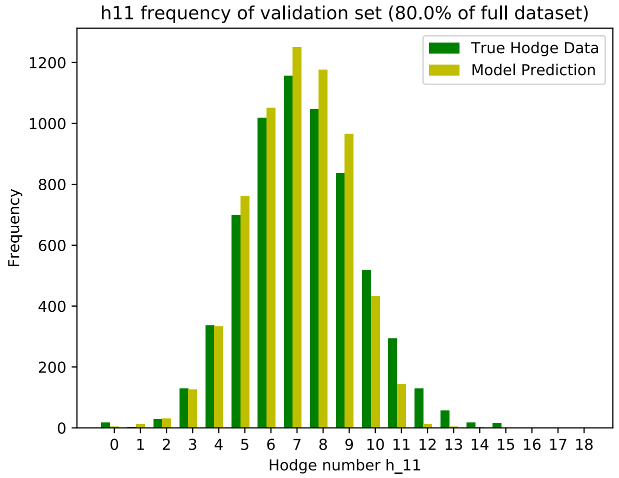
NN classifier



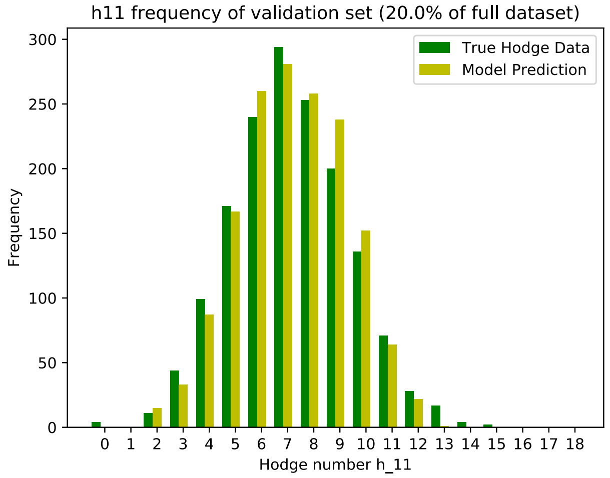


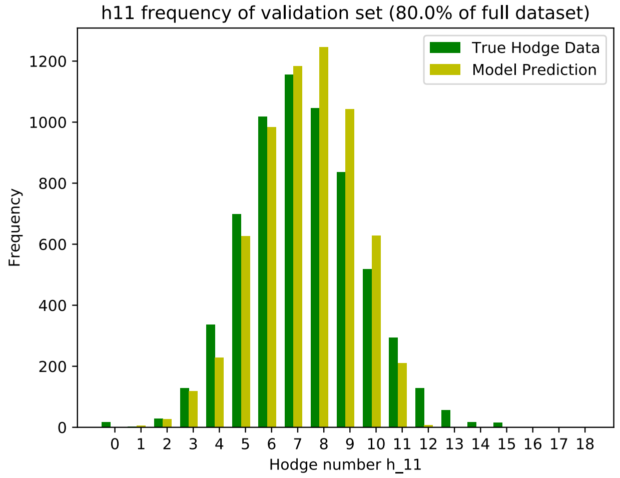
NN regressor





SVM regressor





Bull, He, VJ, Mishra (2018)

Quo Vadis?

The Good	During the last 10-15 years, several international collaborations have computed geometrical and physical quantities and compiled them in vast databases that partially describe the string landscape
The Bad	Computations are hard, especially for a comprehensive treatment: dual cone algorithm (exponential), triangulation (exponential), Gröbner basis (double exponential), how to construct stable bundles over Calabi–Yau manifolds constructed from half a billion polytopes?
The Possibly Beautiful	Borrow techniques from “Big Data”

Machine Learning CICYs

- Subsequent work on topology of CICYs

Bull, He, VJ, Mishra (2019)

Erbin, Finotello (2020)

- Metrics on CICYs

— not known analytically

— needed, *e.g.*, to compute mass of electron

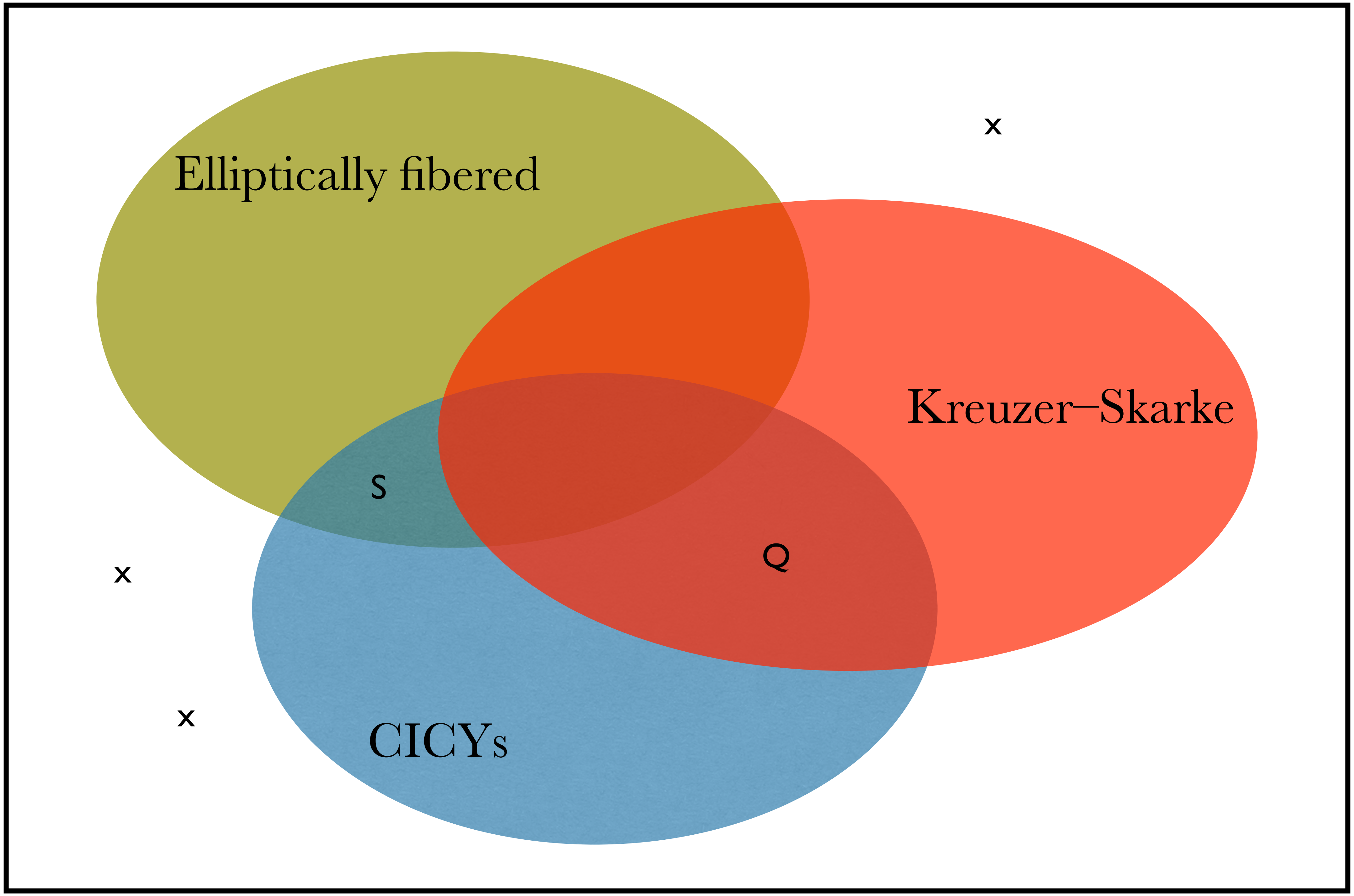
Ashmore, Ovrut, He (2019)

Anderson, Gerdes, Gray, Krippendorf, Raghuram, Rühle (2020)

Douglas, Lakshminarasimhan, Qi (2020)

VJ, Mayorga Peña, Mishra (2020)

Calabi–Yau Threefolds

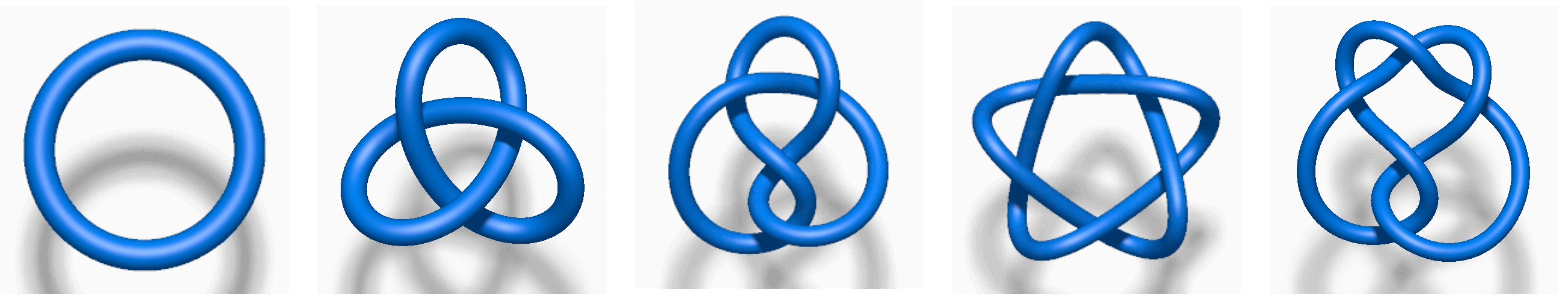


- Reid's fantasy: space of Calabi–Yaus is connected

KNOT THEORY

Dramatis Personae

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



unknot
0_1

trefoil
3_1

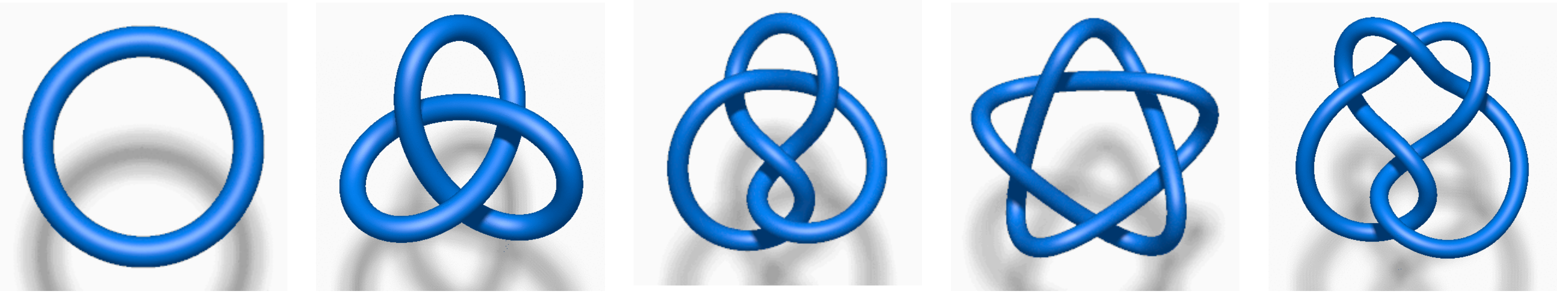
figure-eight
4_1

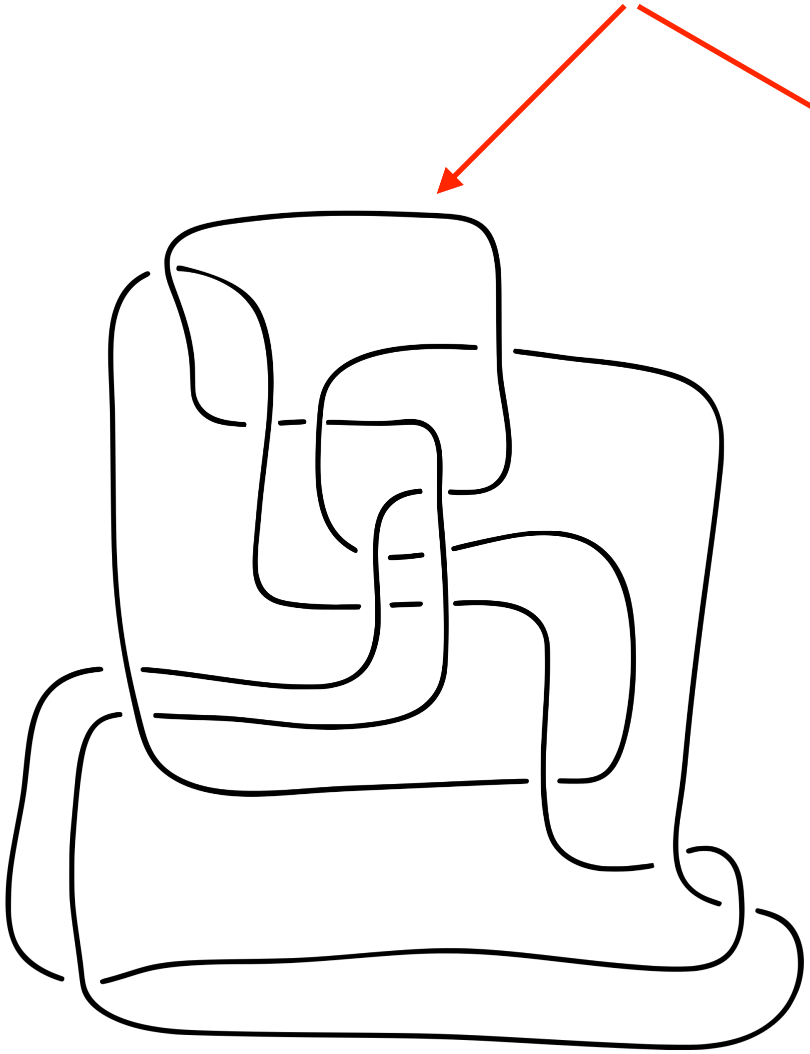
cinquefoil
5_1

three-twist
5_2

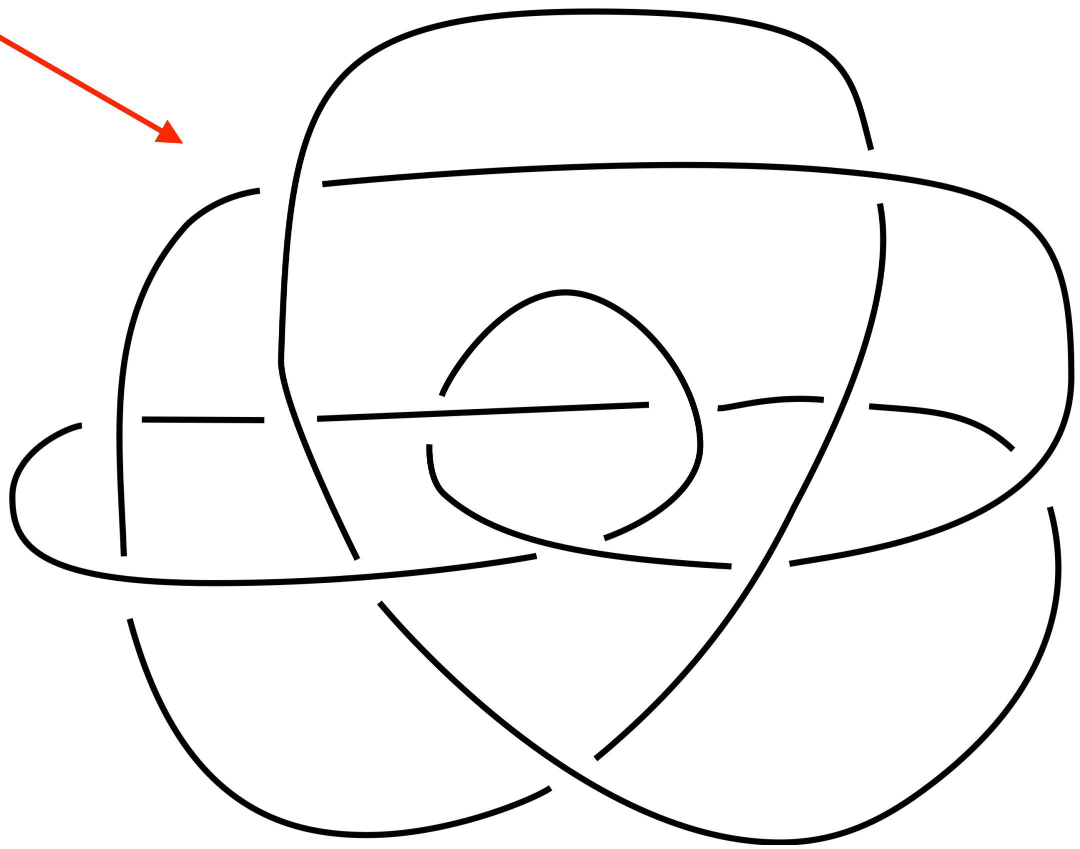
Dramatis Personae

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





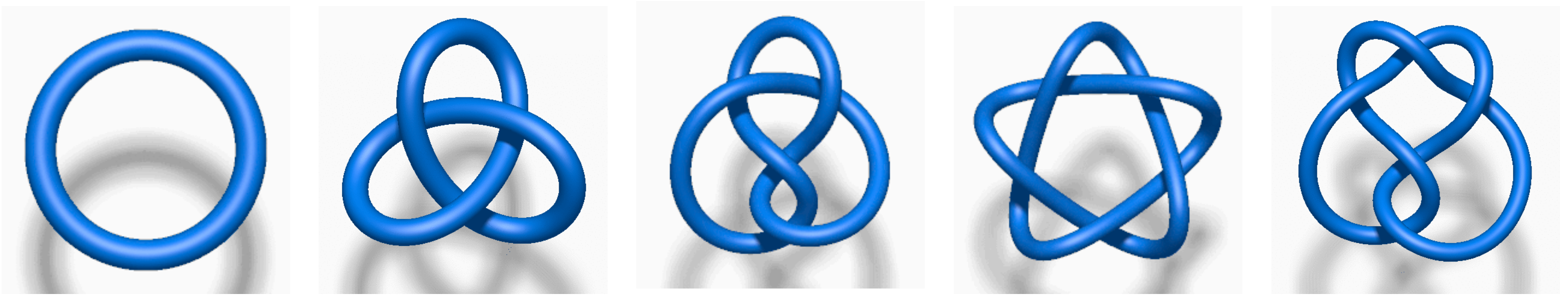
Thistlethwaite unknot



Ochiai unknot

Dramatis Personae

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



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

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

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

$J(\bigcirc;q) = 1$

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

Dramatis Personae

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

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

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

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

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

Jones (1985)
Witten (1989)

$$J_2(4_1;q) = q^{-2} - q^{-1} + 1 - q + q^2 \ , \quad q = e^{\frac{2\pi i}{k+2}}$$

Hyperbolic volume: volume of $S^3 \backslash K$ is another knot invariant computed from tetrahedral decomposition of knot complement

Thurston (1978)
Mostow (1968)

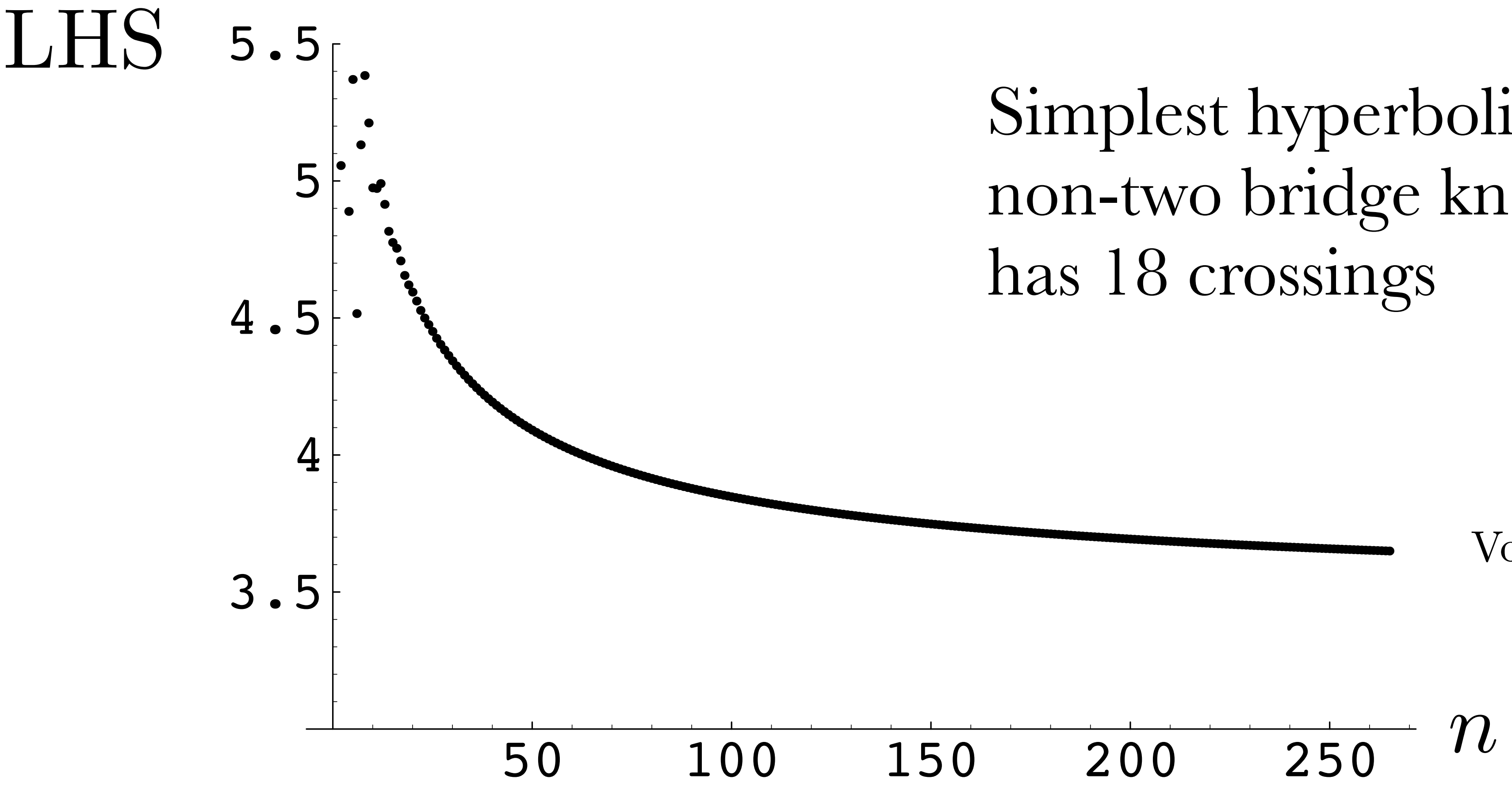
Dramatis Personae

Volume conjecture: $$\lim_{n\to\infty} \frac{2\pi \log |J_n(K;\omega_n)|}{n} = \mathrm{Vol}(S^3 \setminus K)$$

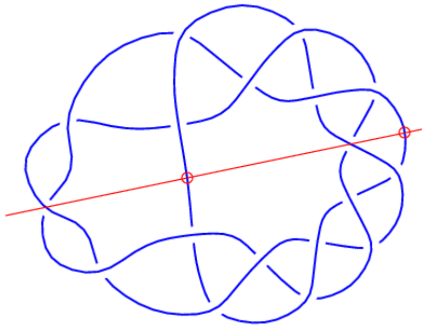
Kashaev (1997)
Murakami x 2 (2001)
Gukov (2005)

$$\omega_n = e^{\frac{2\pi i}{n}}$$

In fact, we take $n, k \to \infty$



Simplest hyperbolic non-two bridge knot, has 18 crossings



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

$\mathrm{Vol}(S^3 \setminus K_0) = 3.474247\ldots$

Behavior is not monotonic!

Garoufalidis, Lan (2004)

Dramatis Personae

Volume conjecture:

$$\lim_{n\to\infty} \frac{2\pi \log |J_n(K;\omega_n)|}{n} = \mathrm{Vol}(S^3 \setminus K)$$

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

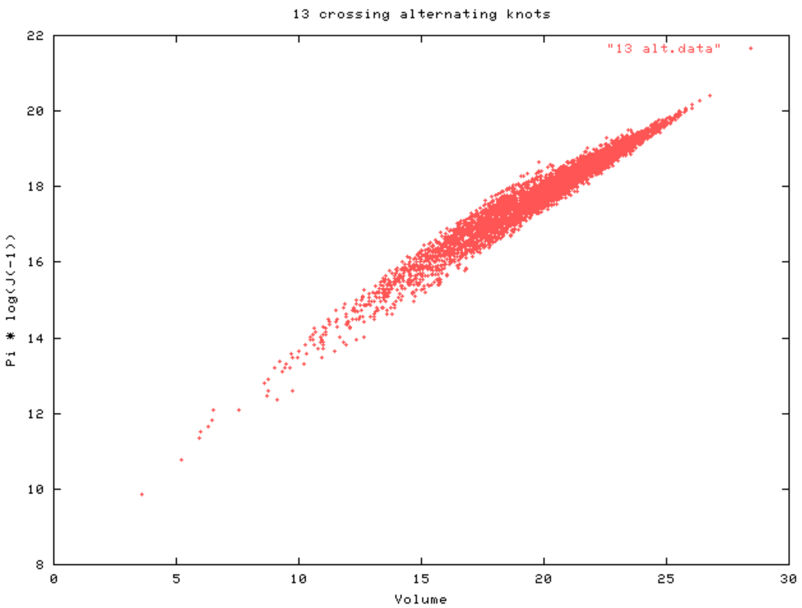
$$\omega_n = e^{\frac{2\pi i}{n}}$$

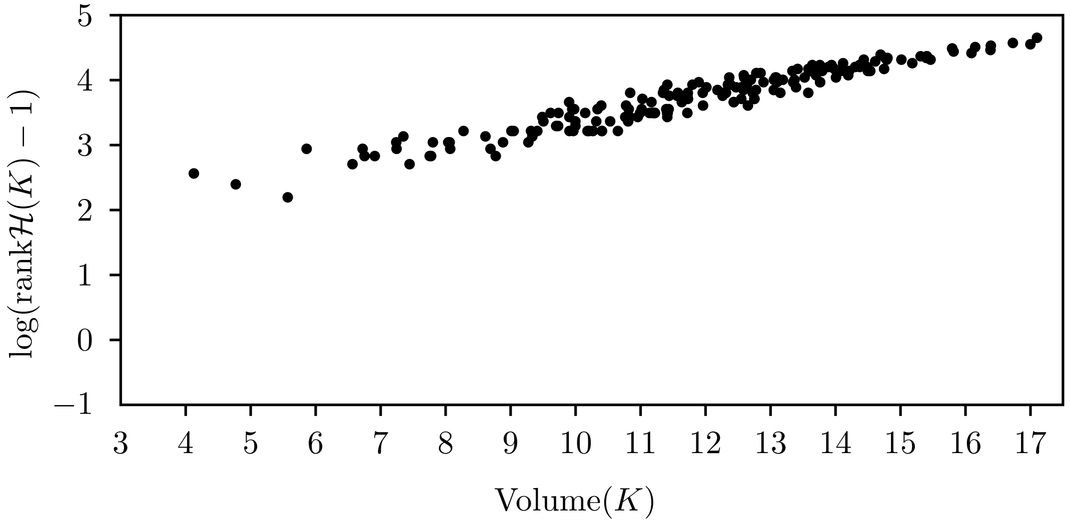
Khovanov homology: a homology theory $\mathcal{H}_K$ whose graded Euler characteristic is $J_2(K;q)$; explains why coefficients are integers

Khovanov (2000)
Bar-Natan (2002)

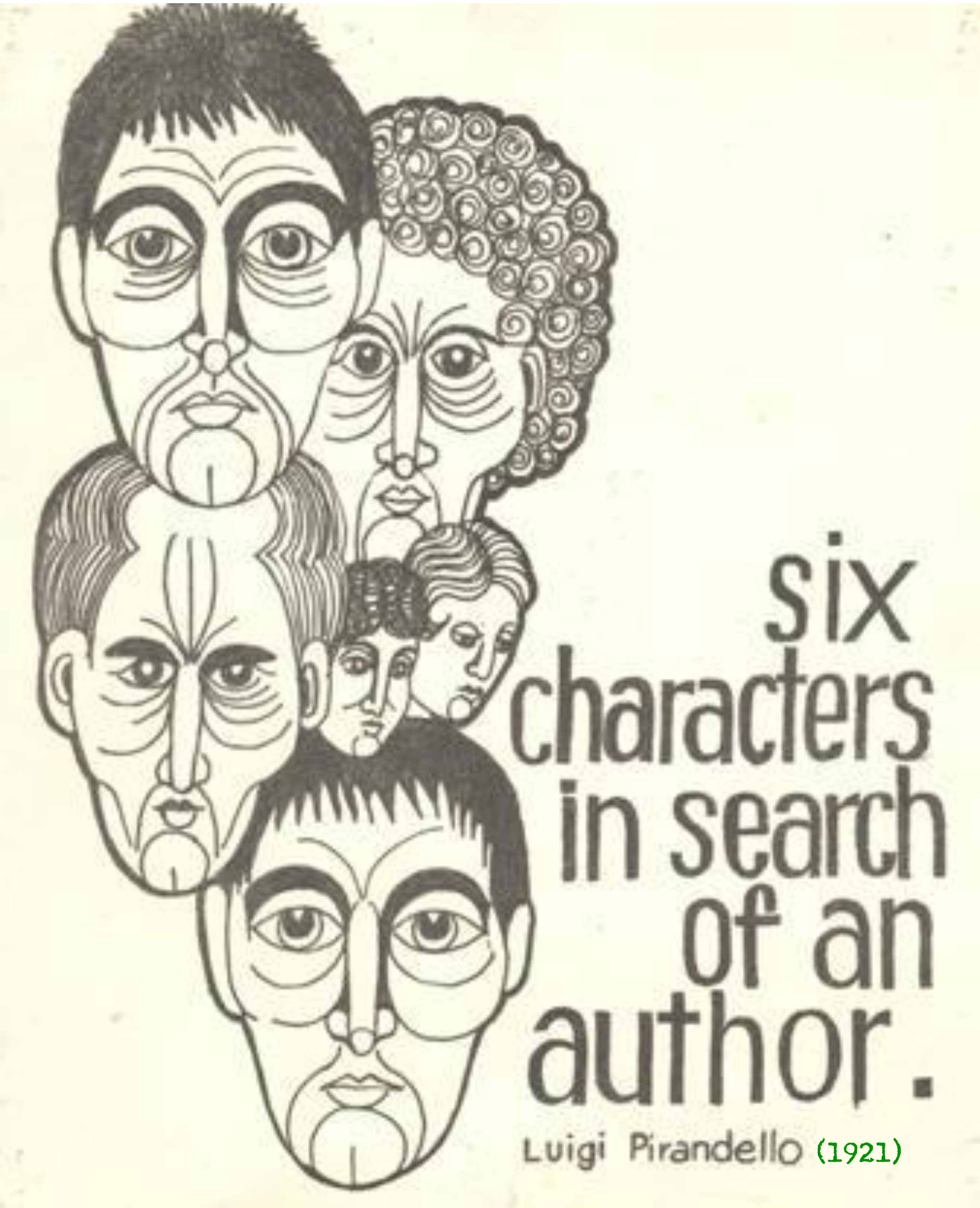
$$\log |J_2(K;-1)| \ , \ \ \log(\mathrm{rank}(\mathcal{H}_K) - 1) \propto \mathrm{Vol}(S^3 \setminus K)$$

Dunfield (2000)
Khovanov (2002)

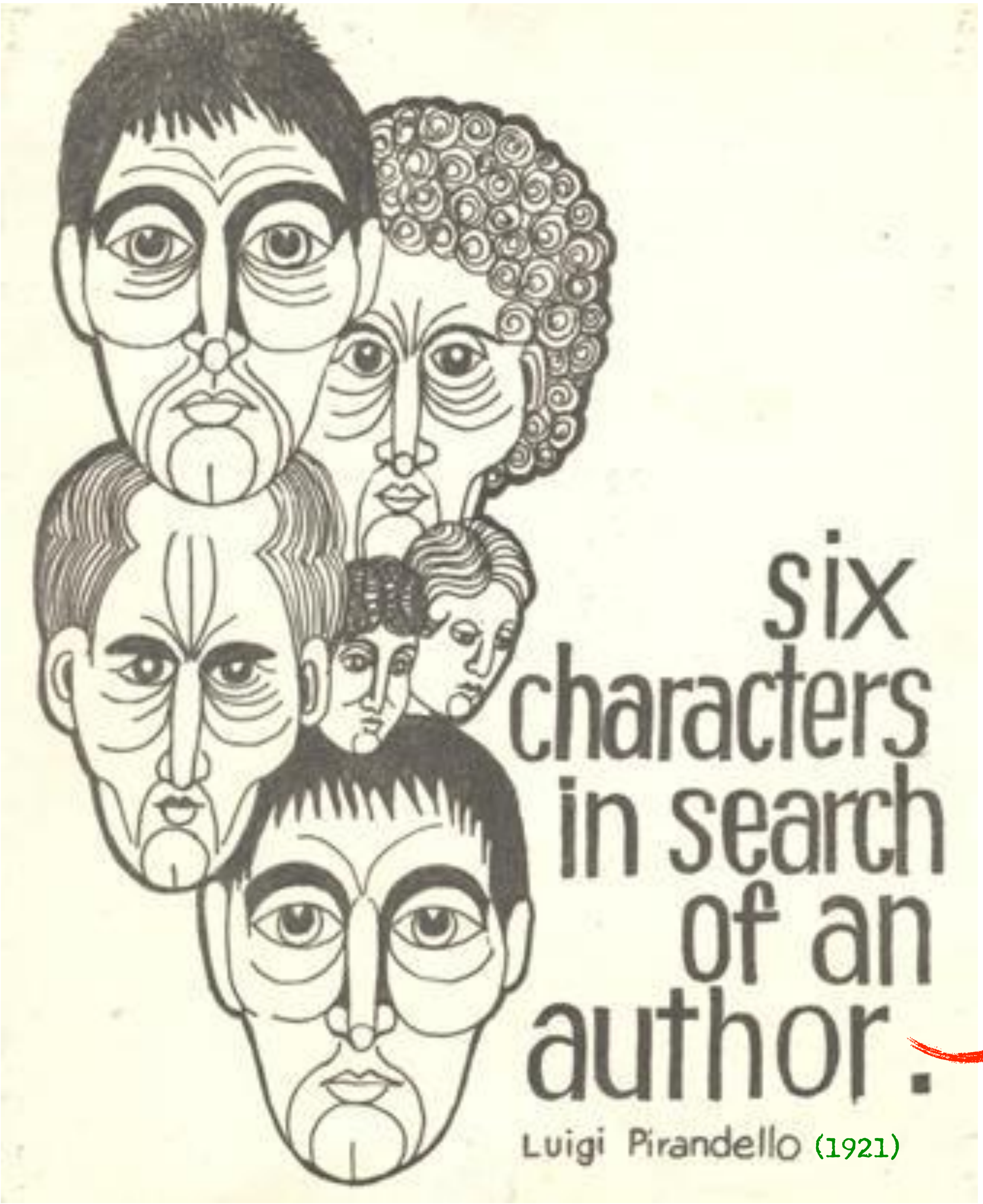




Dramatis Personae

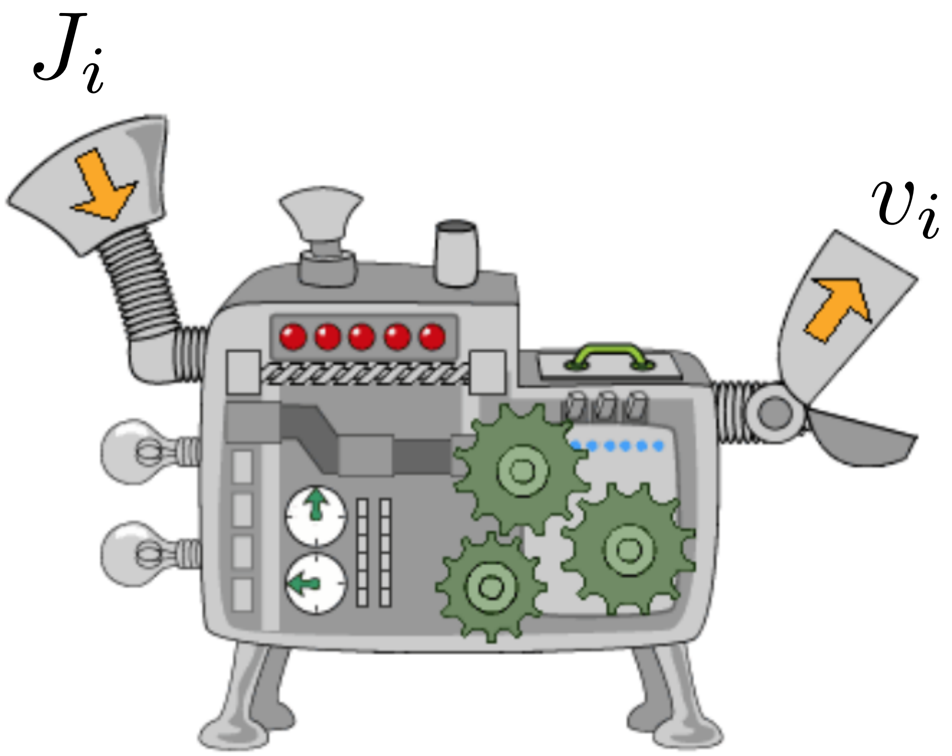


Dramatis Personae



neural network

Neural Network



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

$$J_i \in T$$

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

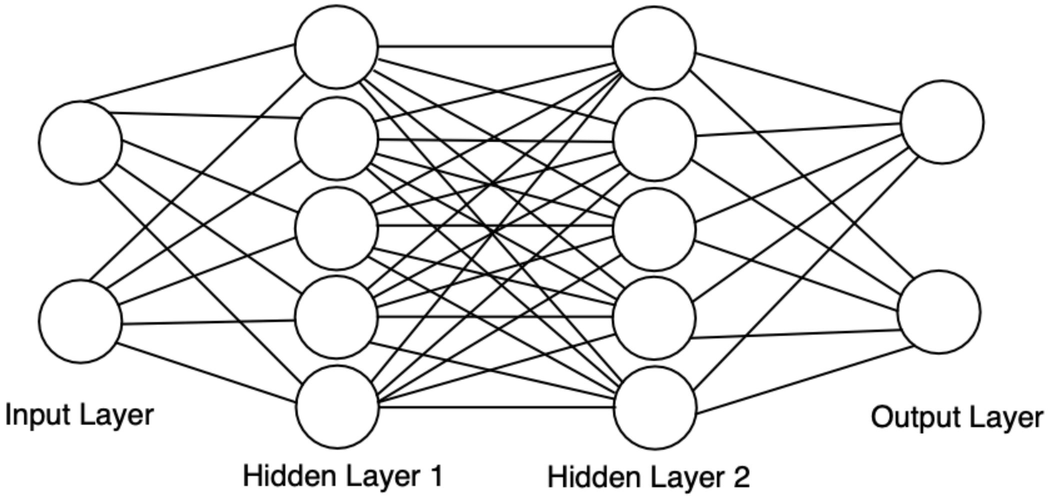
$$J'_i \in T^c$$

Jones polynomials are represented as 18-vectors

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

Two layer neural network in **Mathematica**

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

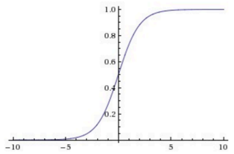


$\vec{J}$ $\quad 100 \times 18 \quad 100 \times 100 \quad \sum_{a=1}^{100}$

12000 hyperparameters

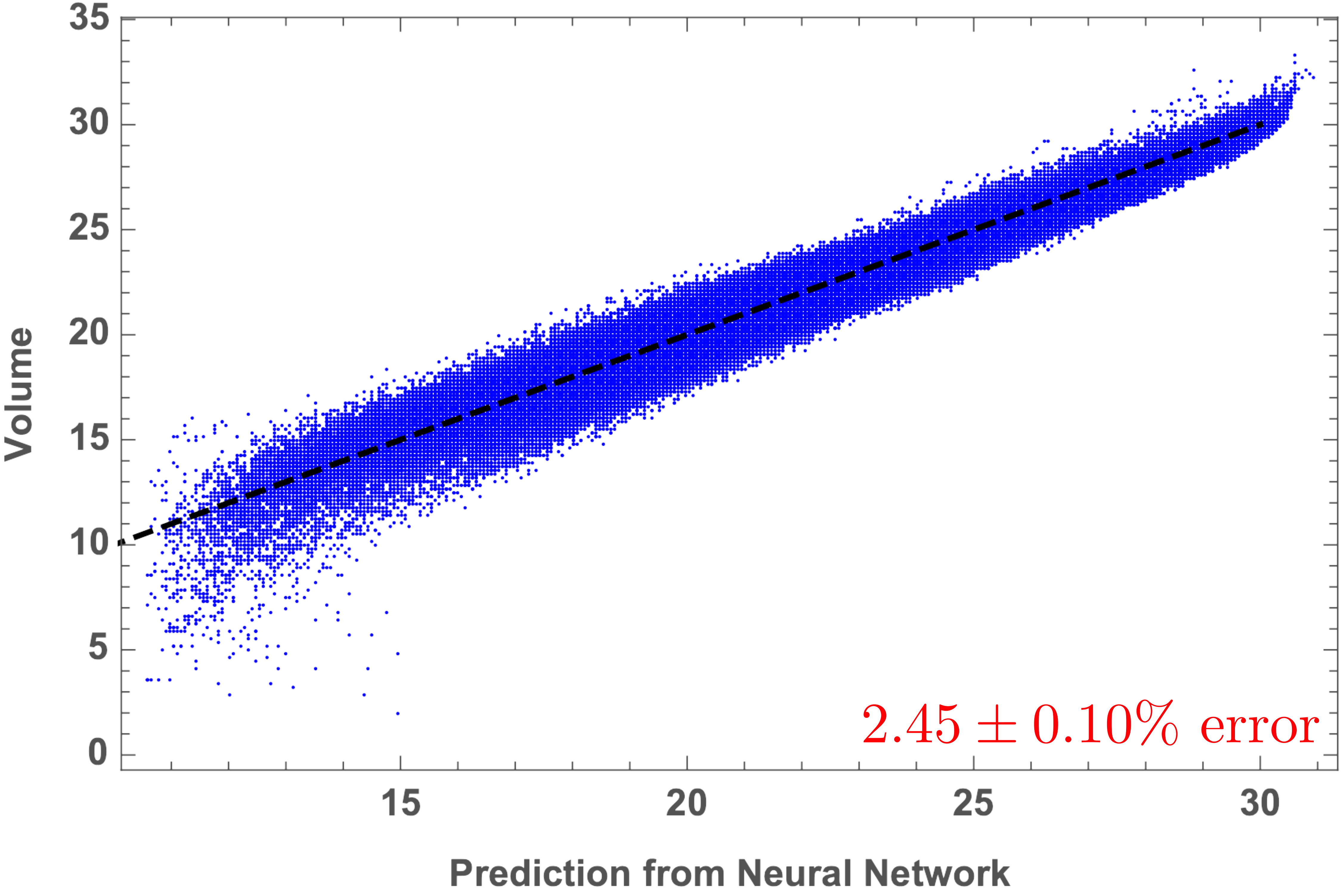
Logistic sigmoids for the hidden layers

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



VJ, Kar, Parrikar (2019)

Neural Network



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

VJ, Kar, Parrikar (2019)

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

- Same applies to 1,701,913 hyperbolic knots up to 16 crossings

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

`VJ, Kar, Parrikar (2019)`

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

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

  weak coupling limit of $SL(2,\mathbb{C})$ Chern–Simons

  strong coupling limit of $SU(2)$

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

$$\lim_{n\to\infty} \frac{2\pi \, \log J_n(K; e^{2\pi i/n})}{n} = \mathrm{Vol}(S^3 \setminus K) + 2\pi^2 i \, \mathrm{CS}(S^3 \setminus K)$$

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

VJ, Kar, Parrikar (2019)

Result

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

- Universal Approximation Theorem: 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)
Hornik (1991)

- Surprise here is simplicity of architecture that does the job

- Ours is in fact the best result in this direction

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

VJ, Kar, Parrikar (2019)

Entr’acte

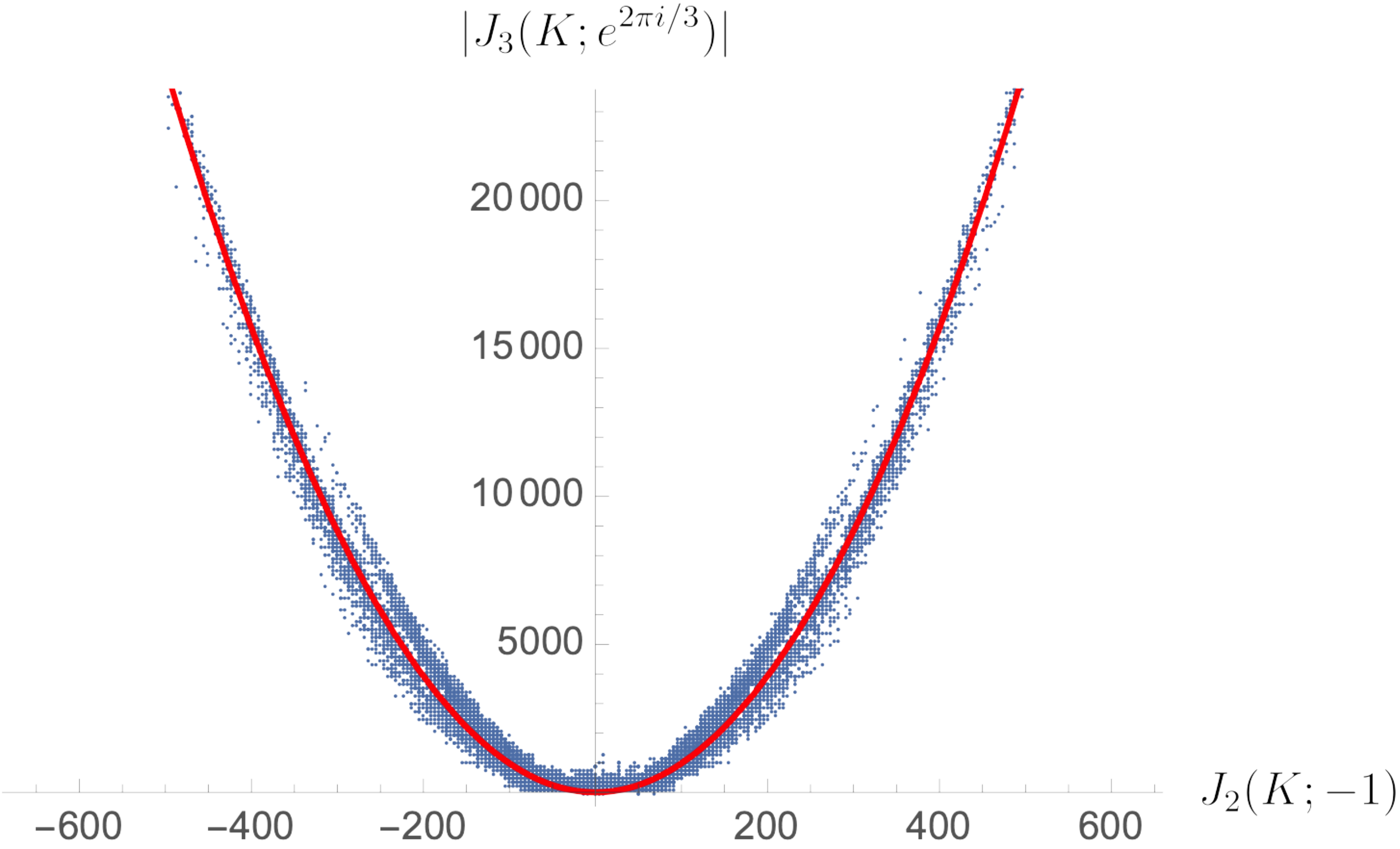
$$v_i = f(J_i) + \text{small corrections}$$

We seek to reverse engineer the neural network to obtain an analytic expression for the volume as a function of the Jones polynomial

To interpret the formula, we use machinery of analytically continued Chern–Simons theory

Towards the Volume Conjecture

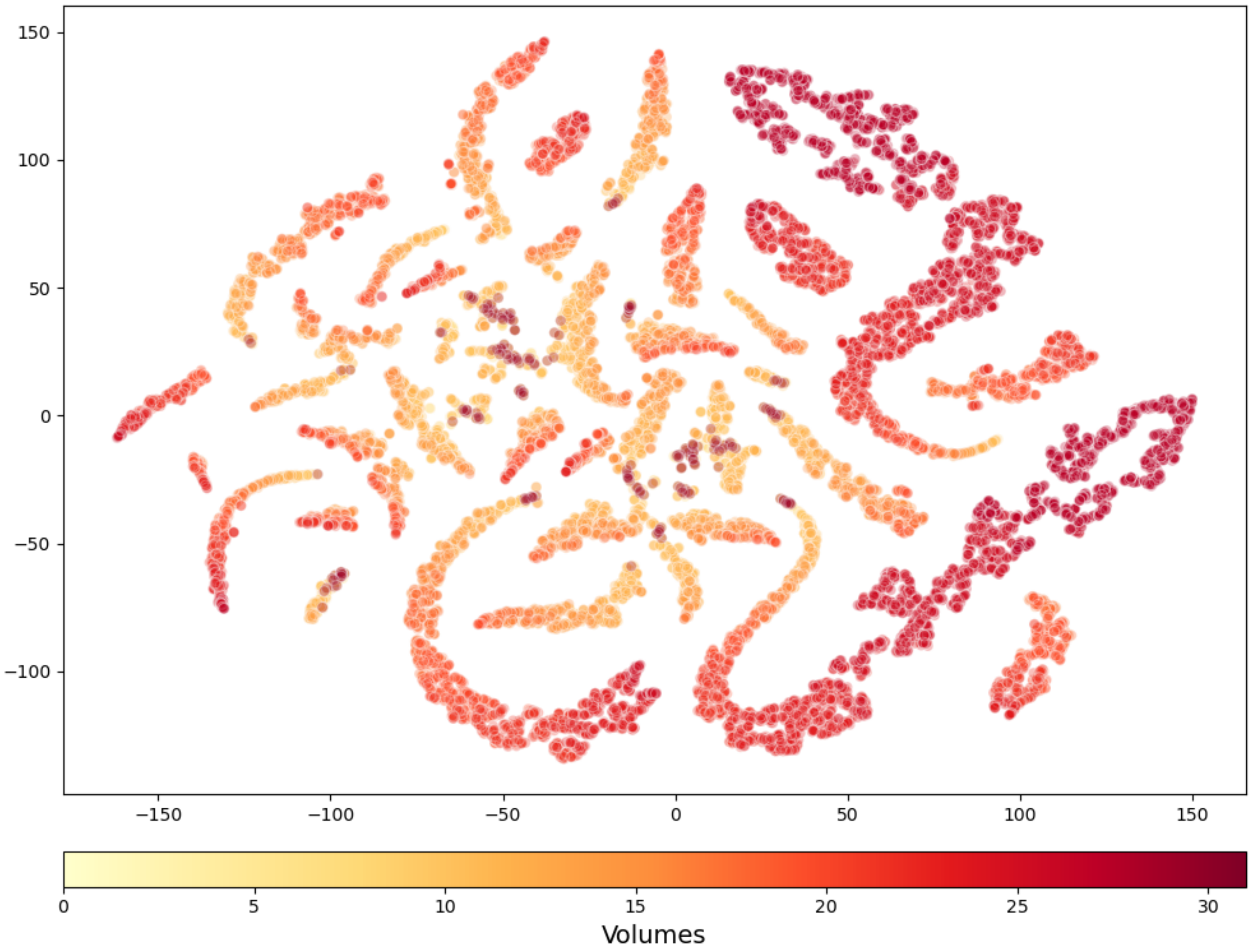
- The volume conjecture: $\lim_{n\to\infty} \frac{2\pi \log |J_n(K;\omega_n)|}{n} = \mathrm{Vol}(S^3 \setminus K)$



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

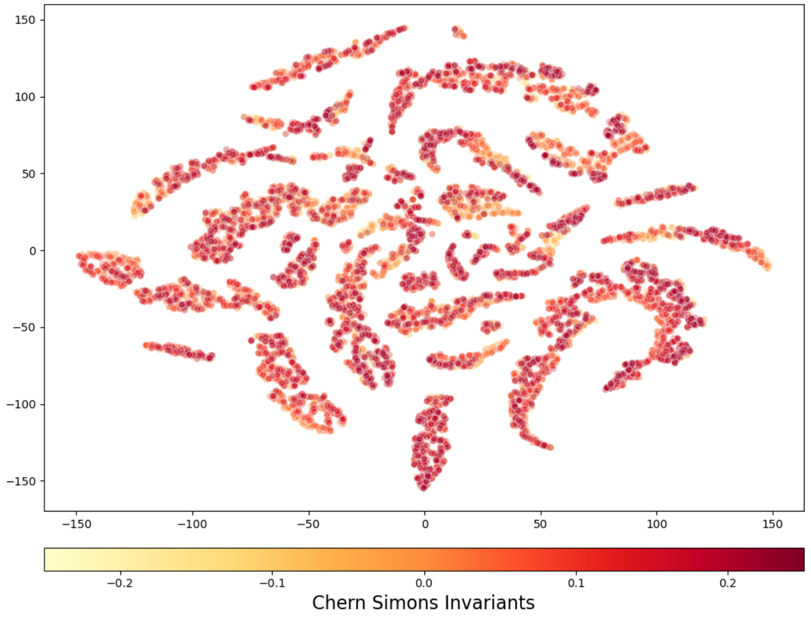
Craven, VJ, Kar (2020)

t-SNE



$$\lim_{n\to\infty} \frac{2\pi \log J_n(K;\omega_n)}{n}$$

$$= \mathrm{Vol}(S^3 \setminus K) + 2\pi^2 i\, \mathrm{CS}(S^3 \setminus K)$$



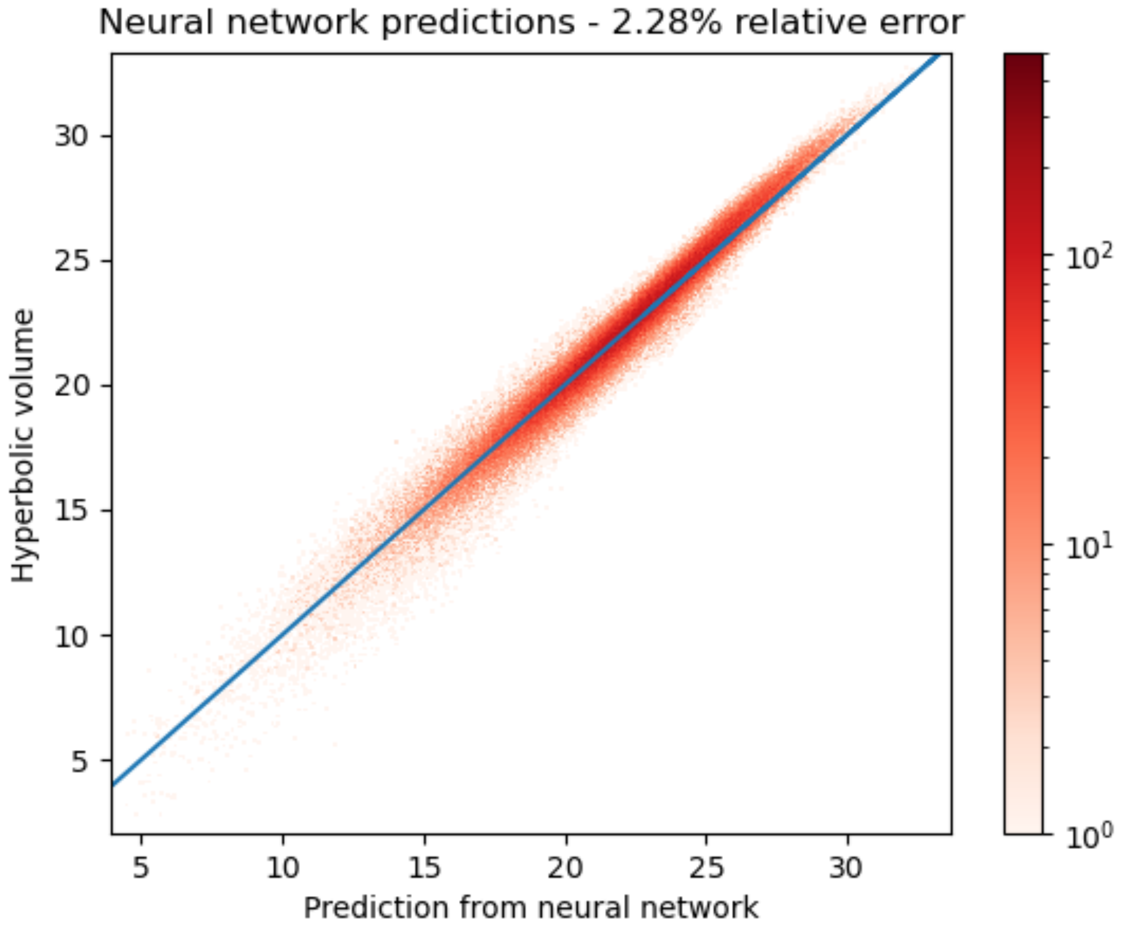
Volume is learnable from coefficients

Chern–Simons invariant probably is not

Craven, VJ, Kar (2020)

No Degrees Needed

- Suppose we drop the degrees and provide only the coefficients; Jones polynomial is no longer recoverable from the input vector
- Results are unchanged!



__N.B.__: we have switched to **Python 3** using **GPU-Tensorflow** with **Keras** wrapper

two hidden layers, 100 neurons/layer, ReLu activation, mean squared loss, **Adam** optimizer

Craven, VJ, Kar (2020)

Jones Evaluations

- Physics in Chern–Simons theory that leads to volume conjecture may also be responsible for information in $J_2(K;q)$

- Consider evaluations of Jones polynomial at roots of unity

- In particular, fix $r \in \mathbb{Z}$ and evaluate $j_p^r := J_2(K; e^{2\pi i p/(r+2)})$

- The input vector

$$\mathbf{v}_{\text{in}} = (\text{Re}(j_0^r), \text{Im}(j_0^r), \ldots, \text{Re}(j^r_{\lfloor (r+2)/2 \rfloor}), \text{Im}(j^r_{\lfloor (r+2)/2 \rfloor}))$$

does not degrade neural network performance

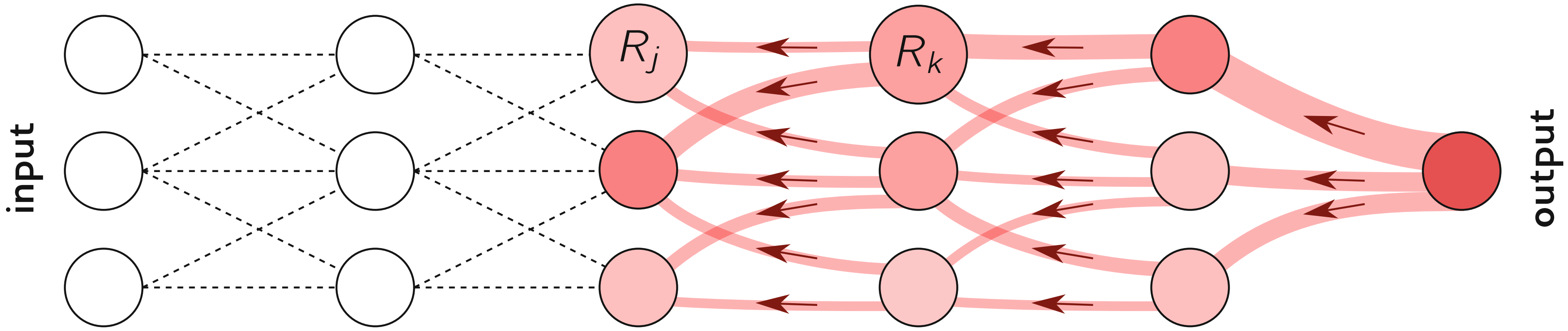
- In fact, we only need to feed in the magnitudes: $\mathbf{v}_{\text{in}} = (|j_0^r|, \ldots, |j_{\lfloor (r+2)/2 \rfloor}|)$

Consistent with degrees not mattering

Craven, VJ, Kar (2020)

Layer-wise Relevance Propagation

- To determine which inputs carry the most weight, propagate backward starting from output layer employing a conservation property



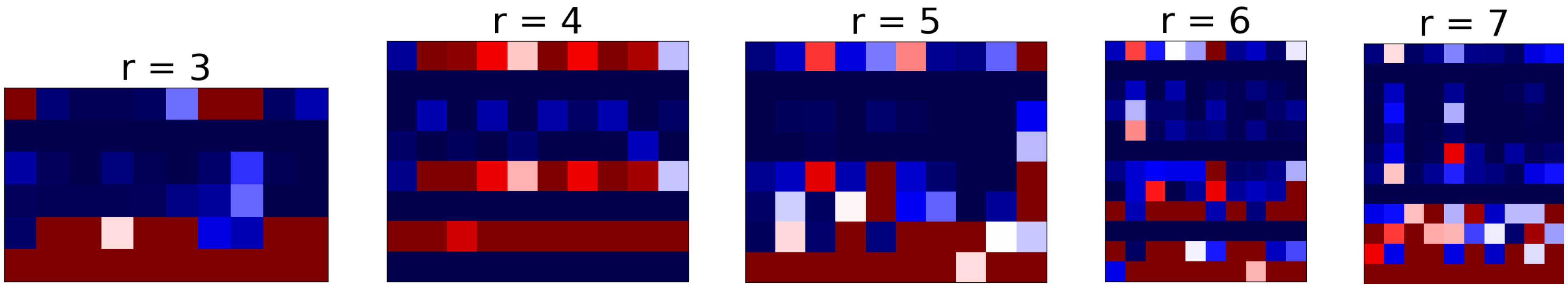
Montavon et al. (2019)

- Compute relevance score for a neuron using activations, weights, and biases

$$R_j^{m-1} = \sum_k \frac{a_j^{m-1} W_{jk}^m + b_k^m}{\sum_l a_l^{m-1} W_{lk}^m + b_k^m} R_k^m \quad , \quad \sum_k R_k^m = 1$$

j^{th} neuron in layer $m - 1$

Layer-wise Relevance Propagation



- Each column is a single input corresponding to evaluations of the Jones polynomial at phases $e^{\frac{2\pi i p}{r+2}}$, $0 \leq 2p \leq r+2$, $p \in \mathbb{Z}$

- Ten different knots

- We show the relevances (red is most relevant) and notice that the same input features light up

Craven, VJ, Kar (2020)

Relevant Phases

r	Error	Relevant roots	Fractional levels	Error (relevant roots)
3	3.48%	$e^{4\pi i/5}$	$\frac{1}{2}$	3.8%
4	6.66%	-1	0	6.78%
5	3.48%	$e^{6\pi i/7}$	$\frac{1}{3}$	3.38%
6	2.94%	$e^{3\pi i/4}, -1$	$\frac{2}{3}, 0$	3%
7	5.37%	$e^{8\pi i/9}$	$\frac{1}{4}$	5.32%
8	2.50%	$e^{3\pi i/5}, e^{4\pi i/5}, -1$	$\frac{4}{3}, \frac{1}{2}, 0$	2.5%
9	2.74%	$e^{8\pi i/11}, e^{10\pi i/11}$	$\frac{3}{4}, \frac{1}{5}$	2.85%
10	3.51%	$e^{2\pi i/3}$,$e^{5\pi i/6}, -1$	$1, \frac{2}{5}, 0$	4.39%
11	2.51%	$e^{8\pi i/13}, e^{10\pi i/13}, e^{12\pi i/13}$	$\frac{5}{4}, \frac{3}{5}, \frac{1}{6}$	2.44%
12	2.39%	$e^{5\pi i/7}, e^{6\pi i/7}, -1$	$\frac{4}{5}, \frac{1}{3}, 0$	2.75%
13	2.52%	$e^{2\pi i/3}, e^{4\pi i/5}, e^{14\pi i/15}$	$1, \frac{1}{2}, \frac{1}{7}$	2.43%
14	2.58%	$e^{3\pi i/4}, e^{7\pi i/8}, -1$	$\frac{2}{3}, \frac{2}{7}, 0$	2.55%
15	2.38%	$e^{12\pi i/17}, e^{14\pi i/17}, e^{16\pi i/17}$	$\frac{5}{6}, \frac{3}{7}, \frac{1}{8}$	2.4%
16	2.57%	$e^{2\pi i/3}, e^{7\pi i/9}, e^{8\pi i/9}, -1$	$1, \frac{4}{7}, \frac{1}{4}, 0$	2.45%
17	2.65%	$e^{14\pi i/19}, e^{16\pi i/19}, e^{18\pi i/19},$	$\frac{5}{7}, \frac{3}{8}, \frac{1}{9}$	2.46%
18	2.49%	$e^{4\pi i/5}, e^{9\pi i/10}, -1$	$\frac{1}{2}, \frac{2}{9}, 0$	2.52%
19	2.45%	$e^{2\pi i/3}, e^{16\pi i/21}, e^{6\pi i/7}, e^{20\pi i/21}$	$1, \frac{5}{8}, \frac{1}{3}, \frac{1}{10}$	2.43%
20	2.79%	$e^{8\pi i/11}, e^{9\pi i/11}, e^{10\pi i/11}, -1$	$\frac{3}{4}, \frac{4}{9}, \frac{1}{5}, 0$	2.4%

$$e^{ix} = e^{\frac{2\pi i}{k+2}}$$

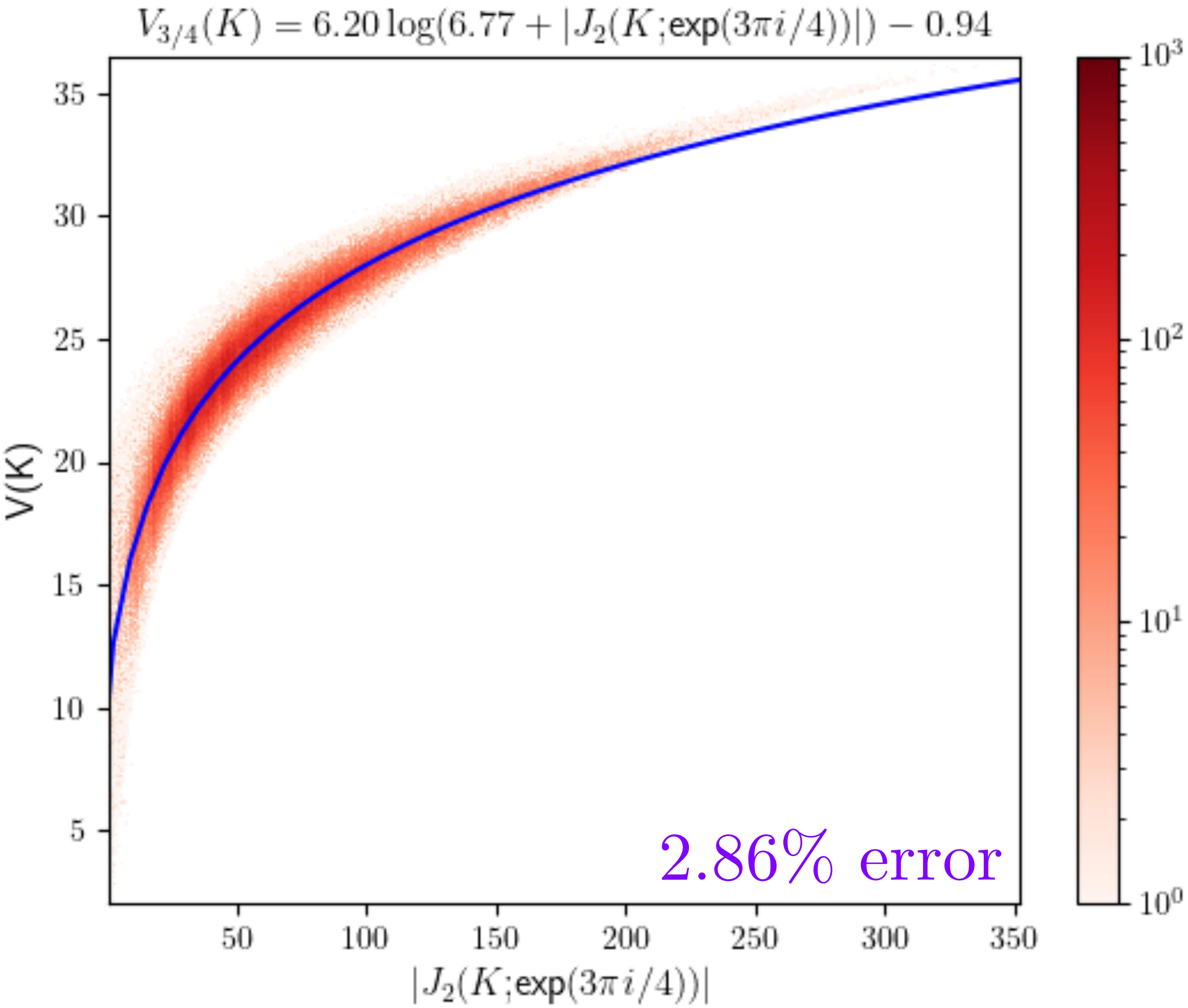
Craven, VJ, Kar (2020)

Phenomenological Function

$$\omega = e^{\frac{3\pi i}{4}}$$

$$\Big\updownarrow$$

$$k = \frac{2}{3}$$



- Parameters fixed via curve fitting routines in **Mathematica**

`Craven, VJ, Kar (2020)`

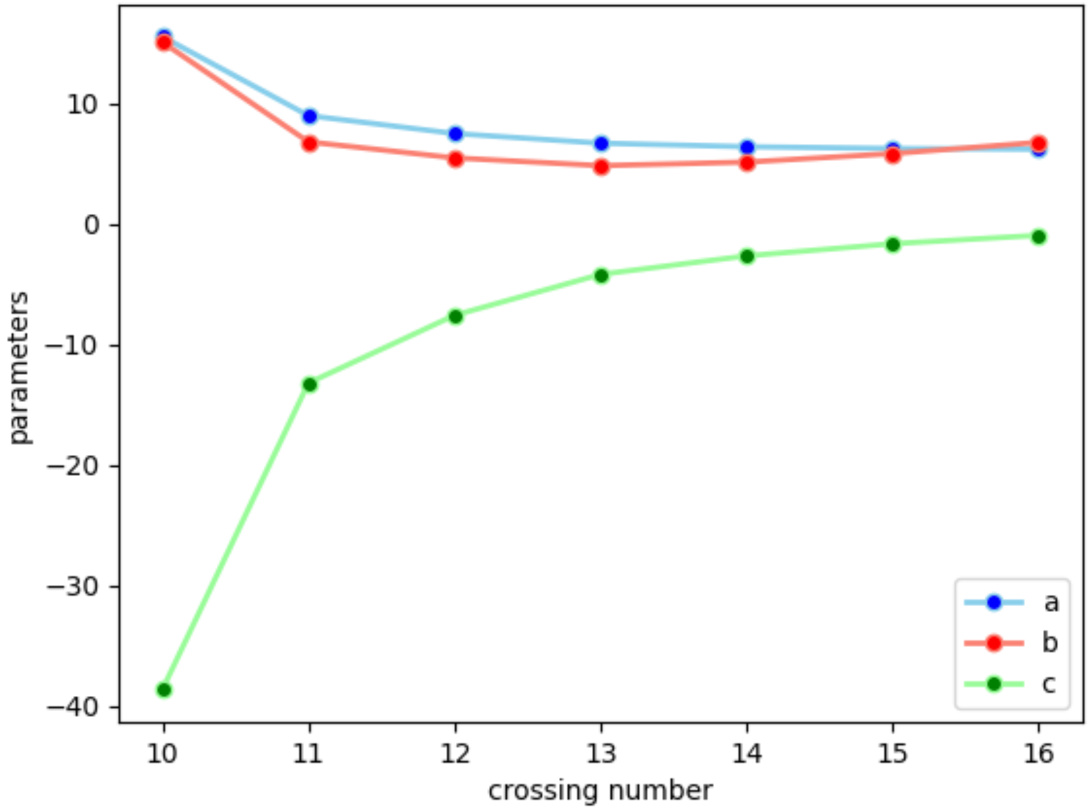
Phenomenological Function

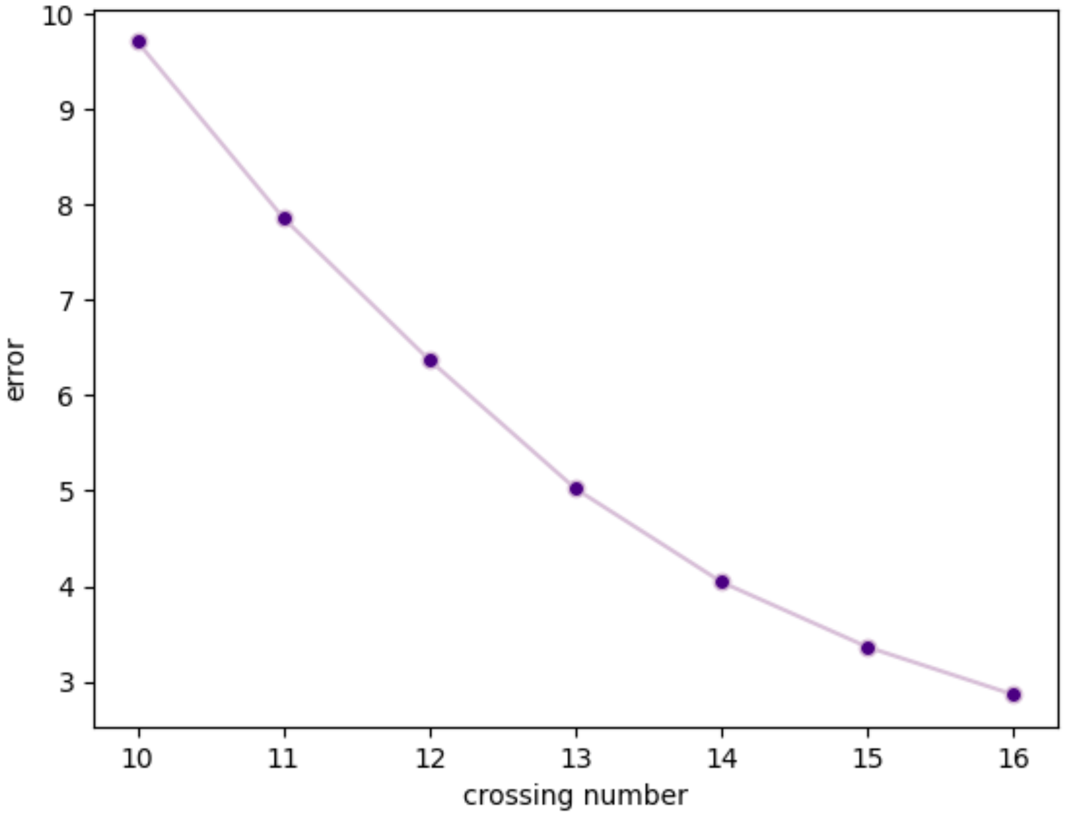
$$\boxed{V_{3/4}(S^3 \setminus K) = 6.20 \log(|J_2(K; e^{\frac{3\pi i}{4}})| + 6.77) - 0.94}$$

2.86% error compared to 2.28% error for neural network

corresponds to Chern–Simons level $k = \frac{2}{3}$

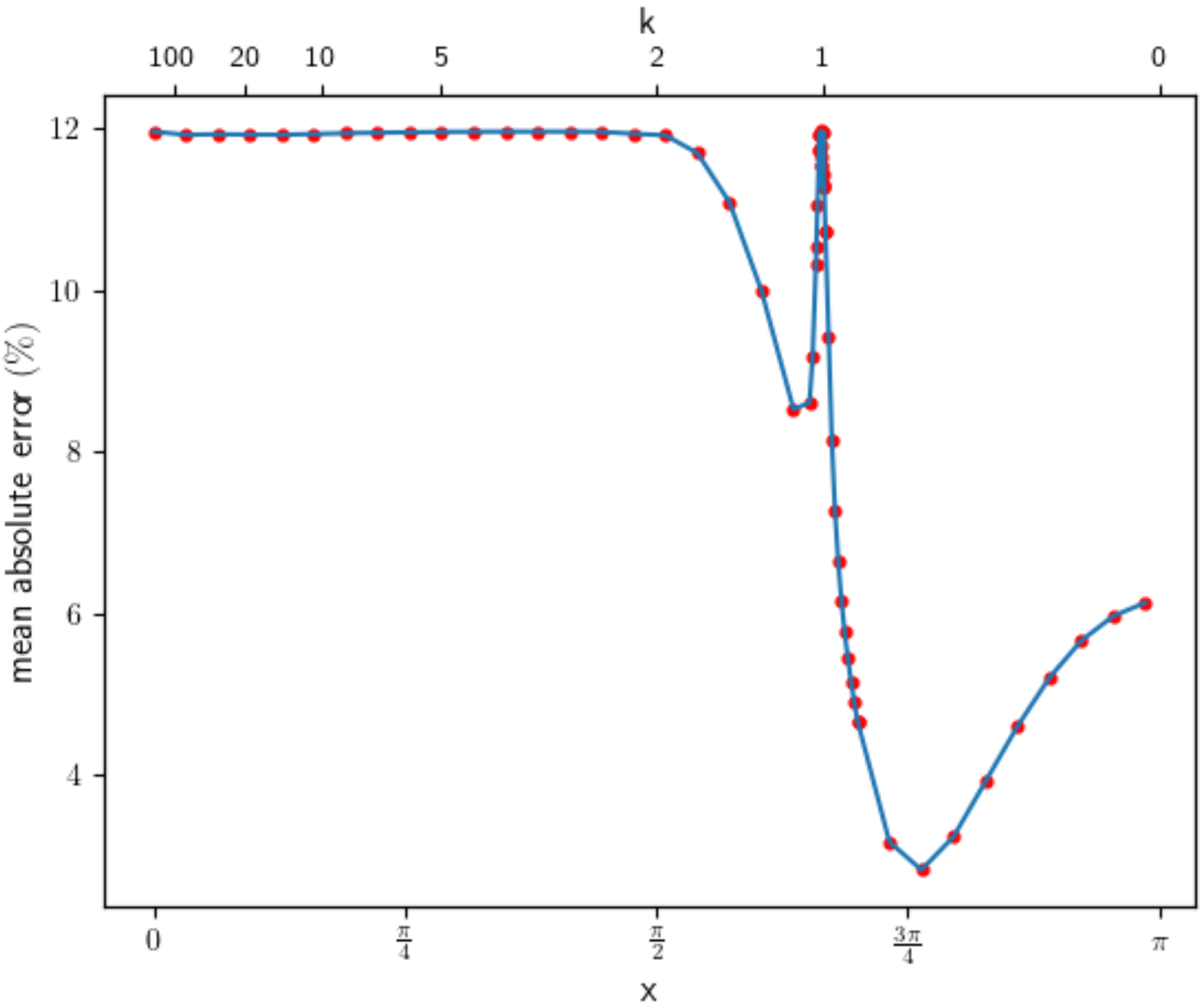
- Parameters of fit robust as a function of crossing number





Craven, VJ, Kar (2020)

The Shape of Things



Craven, VJ, Kar (2020)

A Better Formula

- Our reverse engineered function gave 2.86% error compared to 2.28% error for neural network; the latter is essentially intrinsic

- Can we do better with a formula? If so, how much better?

- Define a new error measure

$$\sigma = \frac{\text{variance of (actual volume − predicted volume)}}{\text{variance of volumes in dataset}}$$

[suggested to us in correspondence with `Fischbacher, Münkler`]

σ-measure is shift/rescaling invariant

- Can ask what fraction of variance is left unexplained

A Better Formula

$$\sigma = \frac{\text{variance of (actual volume − predicted volume)}}{\text{variance of volumes in dataset}}$$

- By this measure, the neural network gives $\sigma = 0.033$

  while our functional approximation gives $\sigma = 0.068$

- If we just assign the average volume to every knot in the dataset, $\sigma = 1$; this corresponds to plateau

- There is room for improvement, but it is remarkable that a function with only three fit parameters works so well

Some Philosophy

The Future

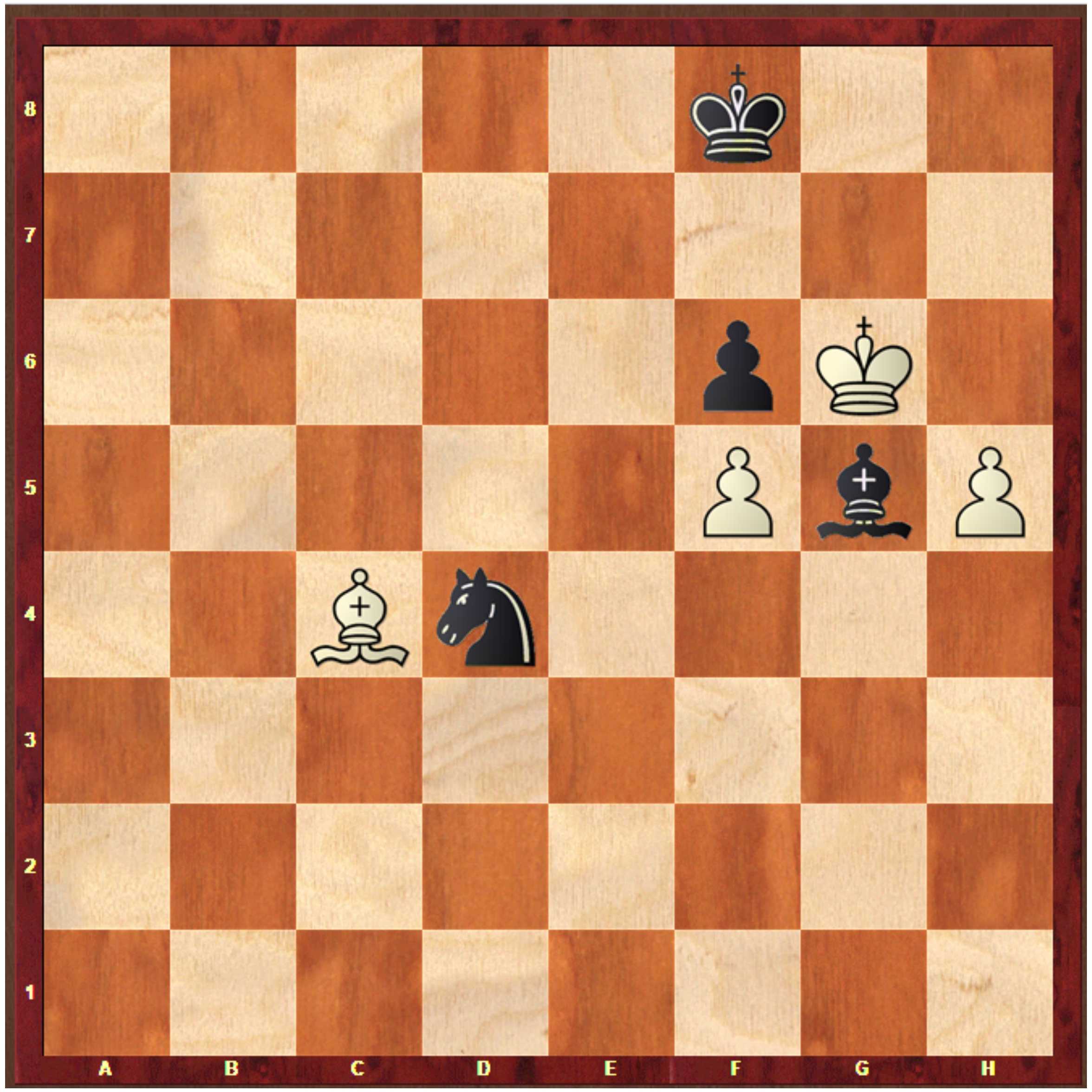
- Machine learning identifies associations

- Want to convert this to analytics — *i.e.*, how does the machine learn?

- What problems in physics and mathematics are machine learnable?

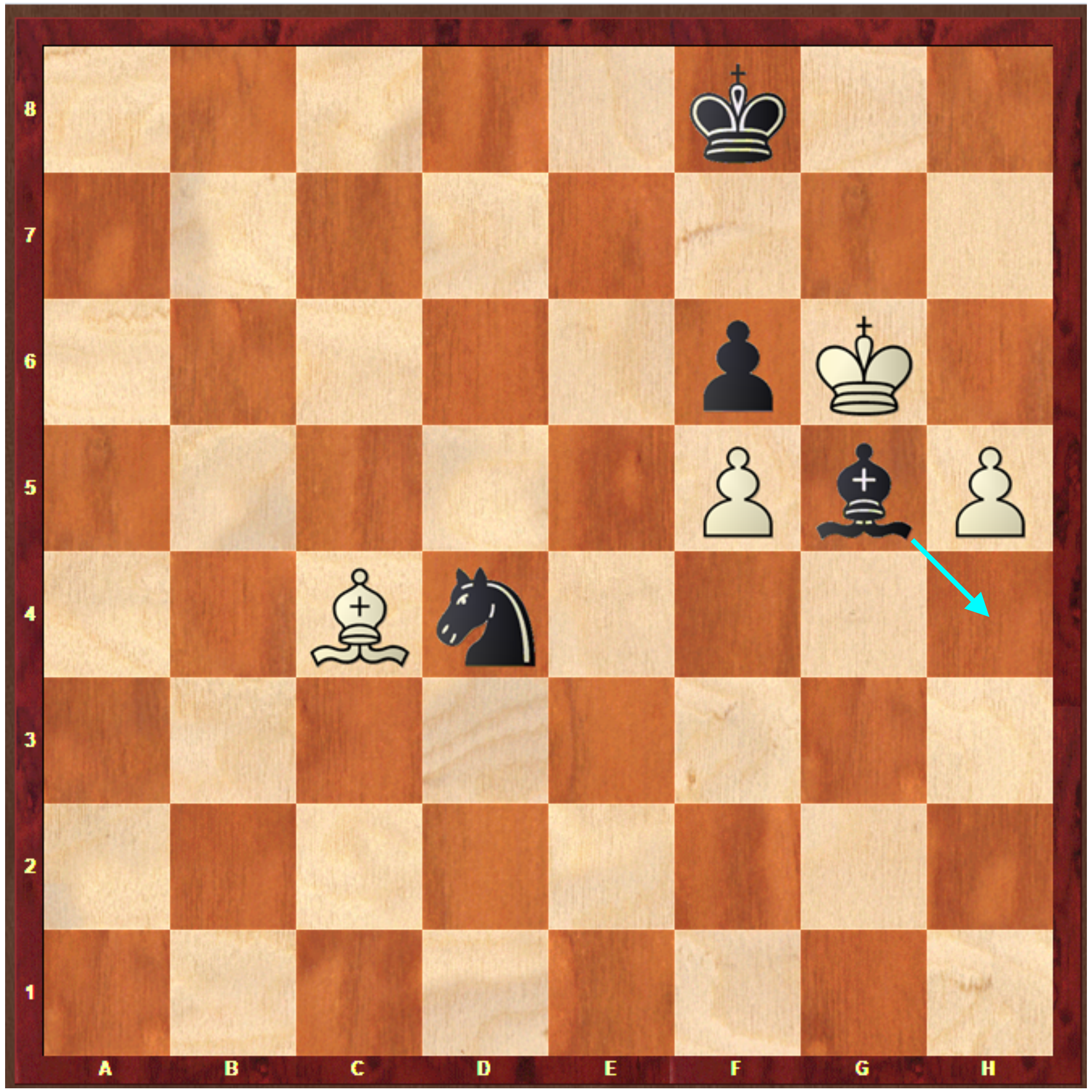
- Can a machine do interesting science?

Stockfish/Sesse



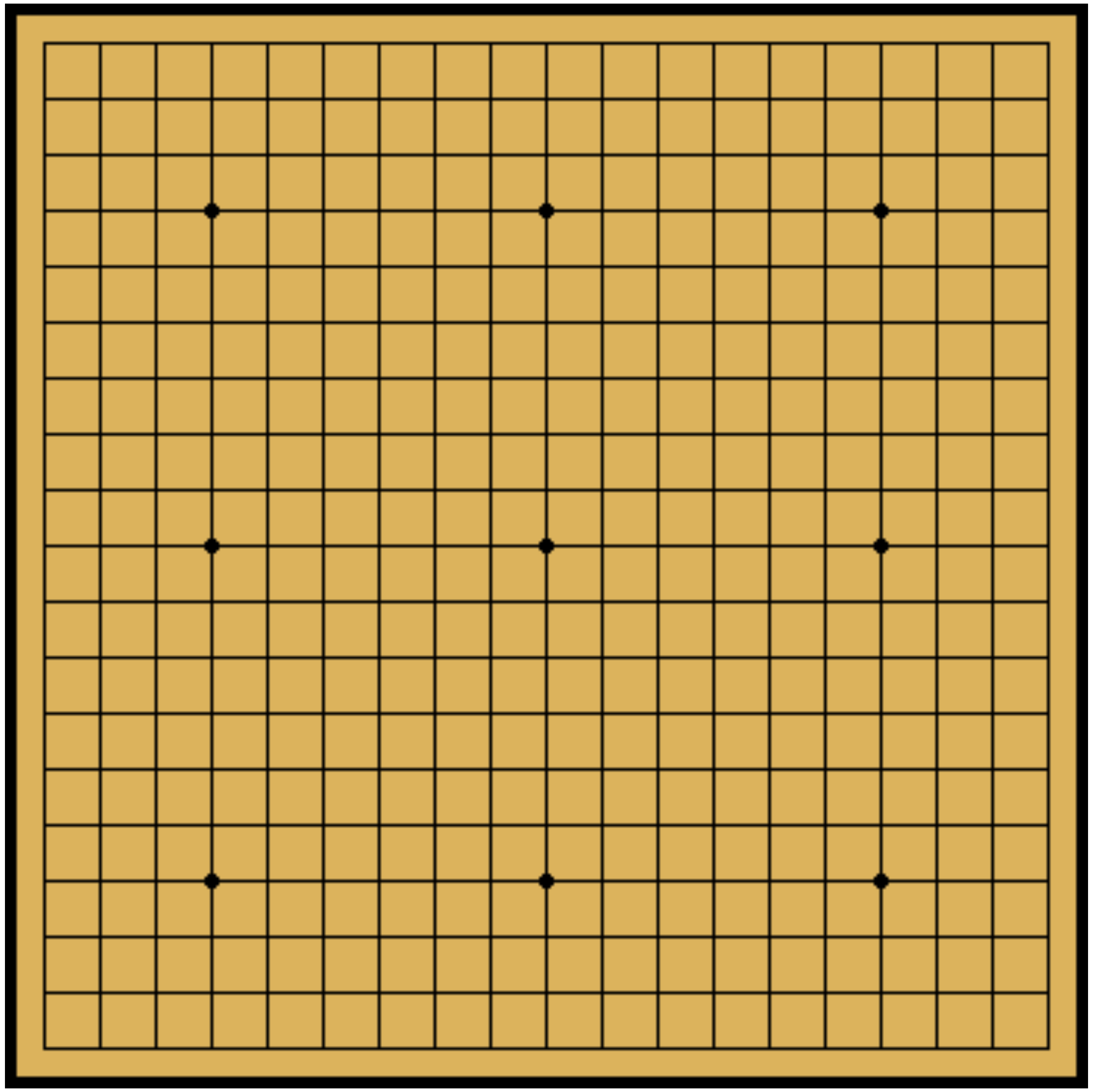
- Carlsen–Caruana, Game 6, World Chess Championship 2018
- Black to move and mate in 36

Stockfish/Sesse



- Carlsen–Caruana, Game 6, World Chess Championship 2018
- Black to move and mate in 36

AlphaZero



- Trained to play Go via self play and it crushes all human players
- Invents new **jōseki**

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 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!