Universes as Bigdata: Geometry, Strings, & AI

YANG-HUI HE

London Institute of Mathematical Sciences, Royal Institution

Merton College, University of Oxford Dept of Mathematics, City, University of London School of Physics, NanKai University

Colloquium: Técnico Lisboa, Jan, 2023

Image: A math the second se

A Classic Problem in Mathematics: Since 1736

- $\bullet\,$ Trichtomy classification of (connected compact orientable) surfaces $\Sigma\,$
 - Euler: topological classification of $\dim_{\mathbb{R}}=2$
 - Gauss: relates topology to metric geometry

Riemann: complexify \rightsquigarrow Riemann surfaces or complex curves: $\dim_{\mathbb{C}} = 1$

		20 8 88	
$g(\Sigma) = 0$	$g(\Sigma)=1$	$g(\Sigma)>1$	
$\chi(\Sigma) = 2$	$\chi(\Sigma)=0$	$\chi(\Sigma) < 0$	
Spherical	Ricci-Flat	Hyperbolic	
+ curvature	$0 \ curvature$	 curvature 	

Euler number $\chi(\Sigma)$, genus $g(\Sigma)$

イロト イヨト イヨト・

$\chi(\Sigma) = 2 - 2g(\Sigma) =$	$= [c_1(\Sigma)] \cdot [\Sigma] =$	$=rac{1}{2\pi}\int_{\Sigma} R =$	$=\sum_{i=0}^{2}(-1)^{i}h^{i}(\Sigma)$
Topology	Algebraic Geometry	Differential Geometry	Index Theorem (co-)Homology
Invariants	Characteristic classes	Curvature	Betti Numbers

イロト イヨト イヨト イ

- $\bullet~\dim_{\mathbb{R}}>2$ manifolds extremely complicated
- Luckily, for a special class of complex manifolds called Kähler

$$g_{\mu\bar{\nu}} = \partial_{\mu}\partial_{\bar{\nu}}K(z,\bar{z})$$

all Σ in $\dim_{\mathbb{C}}=1$ automatically Kähler

• CONJECTURE [E. Calabi, 1954, 1957]: M compact Kähler manifold (g, ω) and $([R] = [c_1(M)])_{H^{1,1}(M)}$.

Then $\exists ! (\tilde{g}, \tilde{\omega})$ such that $([\omega] = [\tilde{\omega}])_{H^2(M;\mathbb{R})}$ and $Ricci(\tilde{\omega}) = R$.

Rmk: $c_1(M) = 0 \Leftrightarrow$ Ricci-flat (rmk: Ricci-flat familiar to physicists through GR)

• THEOREM [S-T Yau, 1977-8; Fields 1982] Existence Proof

< □ > < □ > < □ > < □ > < □ >

Two Pillars of Modern Physics

- MACROSCOPIC (General Relativity)
 - motions of galaxies and stars
 - Gravitational Waves: 2018, new era for cosmology, confidence 99.99994%
- MICROSCOPIC (Quantum Field Theory)
 - $\sim SU(3)_{str} \times [SU(2) \times U(1)]_{EW}$
 - experimentally verified to 19 digits!
 - Higgs Boson: 2014 at LHC, CERN, last piece to SM!





The Greatest Challenge to Science

 Problem: the two are incompatible, unrenormalisability of gravity uncancellable infinities in QFT approach to GR



Standard Model

- Albert Einstein spent the last years of his life on this TOE [theory of everything] in vain "The only thing incomprehensible about our world is that it is comprehensible."
- Should there exist a single equation unifying the Laws of Nature?
- The universe gives a hint:

 $L_{Planck} = \sqrt{\frac{\hbar G_{Newton}}{c^3}} \sim 10^{-35} m$

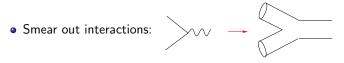


General

Relativity

Paradigm Shift: Points \rightsquigarrow Strings \rightsquigarrow ToE

• Fund constituents 1-dim, NOT points (0-dim)? \sim strings, size L_{Planck}

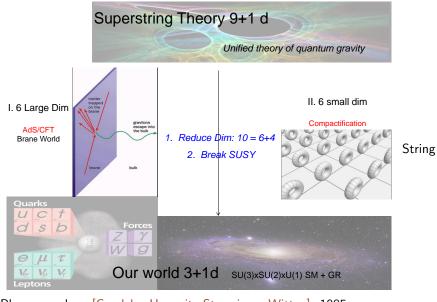


- PREMISE: All particles, all of space-time, all of reality are vibration modes of an OPEN or CLOSED string
 - Heterotic string [Gross-Harvey-Martinec-Rohm]: $E_8 \times E_8$ or SO(32), 1984
 - E_8 accommodates Standard Model

 $SU(3) \times SU(2) \times U(1) \subset SU(5) \subset SO(10) \subset E_6 \subset E_8$

• WORKS only in 10 dimensions

A D F A A F F A



Phenomenology [Candelas-Horowitz-Strominger-Witten]: 1985

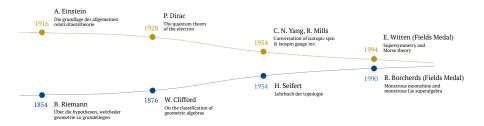
simplest solution of 6 extra dimensions: Ricci-Flat, Kähler $\dim_{\mathbb{C}} = 3$

- Strominger was next door to Yau in 1986 at the IAS, physicists called Ricci-Flat, Kähler manifolds, CHSW called these Calabi-Yau manifolds
- GEOMETRIZATION PROGRAMME: Historically, the right language of physics is increasingly geometrical: (Buy my Books.))
 - Gravity/Space-time \rightsquigarrow GR \rightsquigarrow Differential geometry;
 - Particle physics/Standard Model → Gauge Theory/Yang-Mills → Algebraic geometry (bundles/connections) + group theory (Lie and Finite groups);
 - Condensed matter physics of topological insulators \rightsquigarrow algebraic topology; ...
 - String theory is a brain-child of this tradition
- TAKE-HOME MESSAGE: Whenever physics and maths converge and generate new ideas, the right things are happening

Image: A matching of the second se

The Confluence of Maths and Physics

Physics



Mathematics

YANG-HUI HE	(London/Oxford/Nankai)	
-------------	------------------------	--

The Confluence of Maths and Physics

1959

2010

The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

> EUGENE P. WIGNER Princeton University

Phil. Trans. R. Soc. A (2010) 368, 913-926

Geometry and physics

By Michael Atiyah¹, Robbert Dijkgraaf^{2,*} and Nigel Hitchin³

¹School of Mathematics, University of Edinburgh, Edinburgh EH9 312, UK ²Institute for Theoretical Physics, University of Amsterdam, Volckenierstraat 65, 1018 Amsterdam, The Netherlands ³Mathematical Institute, University of Oxford, 24–29 St Gües, Oxford OXI 3LB, UK

"One may be tempted to invert Wigner's comment and marvel at 'the unreasonable effectiveness of physics in mathematics.'"

Back to Calabi-Yau: Explicit Examples

An interesting sequence: 1,2, ??? ...

$$\dim_{\mathbb{C}} = 1$$

$$\operatorname{Torus} T^{2} = S^{1} \times S^{1}$$

$$\operatorname{QFT} \text{ in } 10 - 2 = 8d$$

$$(1) 4 \operatorname{Torus} T^{4} = S^{1} \times S^{1} \times S^{1} \times S^{1}$$

$$\operatorname{dim}_{\mathbb{C}} = 2$$

$$(2) \text{ K3 surface}$$

$$\operatorname{QFT} \text{ in } 10 - 4 = 6d$$

Unclassified ???

 $\dim_{\mathbb{C}} = 3$ (Yau's Conjecture: Finite Number) Desired QFT in 10 - 6 = 4d



The Inevitability of Algebraic Geometry

- How to construct CY3? Realize as vanishing locus of polynomials, Algebraic Geometry e.g., $\{(p,q)|p^2 + q^2 - 1 = 0\} \subset \mathbb{R}^2$ is a circle (1-real dimension)
- Complexify and Projectivize (Projective algebraic variety)
 - Cubic equation in \mathbb{CP}^2 : e.g. CY1 = T^2 { $(x, y, z)|x^3 + y^3 + z^3 = 0$ } $\subset \mathbb{CP}^2$ (elliptic curve); dim_C = 2 - 1 = 1
 - TMH: Homogeneous Eq in \mathbb{CP}^n , degree = n+1 is Calabi-Yau of $\dim_{\mathbb{C}} = n-1$
- An Early Physical Challenge to Algebraic Geometry
 - Particle content in [CHSW] # gens of SM particles $=\pm \frac{1}{2}$ Euler number
 - 1986 Question: Are there Calabi-Yau-3 with $\chi = \pm 6$?

イロト イヨト イヨト イヨト

The First Data-sets in Mathematical Physics/Geometry

- [Candelas-A. He-Hübsch-Lutken-Schimmrigk-Berglund] (1986-1990)
 - CICYs (complete intersection CYs) multi-deg polys in products of \mathbb{CP}^{n_i}
 - Problem: *classify all configuration matrices*; employed the best computers at the time (**CERN supercomputer**); q.v. magnetic tape and dot-matrix printout in Philip's office
 - 7890 matrices, 266 Hodge pairs $(h^{1,1}, h^{2,1})$, 70 Euler $\chi \in [-200, 0]$
- [Candelas-Lynker-Schimmrigk, 1990]
 - Hypersurfaces in Weighted P4
 - 7555 inequivalent 5-vectors w_i , 2780 Hodge pairs, $\chi \in [-960, 960]$
- [Kreuzer-Skarke, mid-1990s 2000]
 - Hypersurfaces in (Reflexive, Gorenstein Fano) Toric 4-folds
 - 6-month running time on dual Pentium SGI machine
 - at least 473,800,776, with 30,108 distinct Hodge pairs, $\chi \in [-960,960]$

Technically, Moses



was the first person with a tablet downloading data from the cloud The age of data science in mathematical physics/string theory not as recent as you might think After 40 years of research by mathematicians and physicists

.

Geometric Origin of our Universe

- Each CY3 (+ bundles, discrete symmetries) X gives a 4-D universe
 - The geometry (algebraic/differential geometry/topology etc.) of X determines the physical properties of the 4-D world
 - particles and interactions \sim cohomology theory; masses \sim metric; Yukawa \sim

Triple intersections/integral of forms over X



Ubi materia, ibi geometria

- Johannes Kepler (1571-1630)
- Our Universe: {

 (1) probabilistic/anthropic?
 (2) Sui generis/selection rule? Triadophilia
 (3) one of multi-verse ?
 cf. Exo-planet/Habitable Zone search (so far ~ 10³ in 10²¹)

۰

Lecture Notes in Mathematics 2293

Yang-Hui He

The Calabi—Yau Landscape

From Geometry, to Physics, to Machine Learning YH He Springer-Nature, 2021

Lecture Notes in Maths

A playful intro for Masters to PhD students

《曰》 《聞》 《臣》 《臣》 三臣

An even more playful intro



Perhaps the biggest theoretical challenge to string theory:

selection criterion ??? metric on the landscape???

- Douglas (2003): Statistics of String vacua
- Kachru-Kallosh-Linde-Trivedi (2003): type II/CY estimates of 10^{500}
- Taylor-YN Wang (2015-7): F-theory estimates 10^{3000} to 10^{10^5}
- Basic Reason:

Algebraic Geometry \rightsquigarrow Combinatorial Geometry \rightsquigarrow Exponential Growth in dim

< □ > < 同 > < 回 > < Ξ > < Ξ

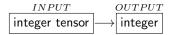
- The Good Last 10-15 years: several international groups have bitten the bullet
 Oxford, London, Vienna, Blacksburg, Boston, Johannesburg, Munich, ... computed
 many geometrical/physical quantities and compiled them into
 various databases Landscape Data (10^{9~10} entries typically)
 The Bad Generic computation HARD: dual cone algorithm (exponential),
 - triangulation (exponential), Gröbner basis (double-exponential), ...e.g., how to construct stable bundles over the $\gg 473$ million KS CY3? Sifting through for SM computationally impossible ...

The ??? Borrow new techniques from "Big Data" revolution

イロト イヨト イヨト イヨ

A Wild Question

• Typical Problem in String Theory/Algebraic Geometry:



- Q: Can problems in computational geometry and theoretical physics be "learned" by AI ? implications:
 - can we "machine-learn the landscape?"
 - can we do mathematics with ML?
- [YHH 1706.02714] Deep-Learning the Landscape, PLB 774, 2017

Science feature article, Aug, vol 365 issue 6452 :

Experimentally, it seems to be the case for many situations in geometry and beyond in a matter of seconds on ordinary laptop.

(日)

Algebraic Geometry is an Image Processing Problem

- Every manifold can be represented as a numerical tensor ۰
- A typical calculation:

$$\rightarrow 22$$

< □ > < □ > < □ > < □ > < □ >

- NN doesn't know/care about the maths Proper Way
- Feed to some Neural Network:
 - $\left\{\begin{array}{l} -\text{ Take samples of } X \to h \text{ (compiled over 40 years)} \\ -\text{ train a NN, or SVM, etc.,} \\ -\text{ Validation on unseen } X \to h \end{array}\right.$
- can reach reaches 99.9% quickly (cf. YHH, Bull-YHH-Jejjala-Mishra,

Erbin-Finotello, Constantin-Lukas, ...)

2017: String Theory enters the ML Era

YHH (1706.02714); (see also: Krefl-Seong (1706.03346); Ruehle (1706.07024) Carifio-Halverson-Krioukov-Nelson (1707.00655))



Sophia: Hanson Robotics,

HongKong

< □ > < □ > < □ > < □ > < □ >

- Beginning of String_Data annual conference series
- How can ML and modern data-science help with the vacuum degeneracy problem & string phenomenology??
- Meanwhile ... Sophia becomes a "human" citizen (in Saudi Arabia)

Machine Learning Mathematics

Why stop at string/geometry?

How do different branches of mathematics respond to ML?

Review Paper: YHH 2101.06317 "Machine-Learning Mathematical Structures", IJDSMS 2022

How does one *DO* mathematics, I ?

- Russell-Whitehead Principia Mathematica [1910s] (Leibniz, Frege, ...) axiomatize maths, but ... Gödel [1931] Incompleteness; Church-Turing [1930s] Undecidability
- Automated Theorem Proving (ATP) "The practicing mathematician hardly ever worries about Gödel"
 - Newell-Simon-Shaw [1956] Logical Theory Machine: subset of Principia
 - Type Theory [1970s] Martin-Löf, Coquand, ... Coq: 4-color (2005);
 Feit-Thompson Thm (2012); Lean (2013); Univalent Foundation / Homotopy
 Type Theory [2006-] Voevodsky

Buzzard: "Future of Maths" 2019, ICM 2022 Davenport: ICM 2018

"Computer Assisted Proofs" Szegedy: more extreme view, computers >

humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

We can call this Bottom-up Mathematics

イロト イポト イヨト イヨト

How does one *DO* mathematics, I ?

- Russell-Whitehead Principia Mathematica [1910s] (Leibniz, Frege, ...) axiomatize maths, but ... Gödel [1931] Incompleteness; Church-Turing [1930s] Undecidability
- Automated Theorem Proving (ATP) "The practicing mathematician hardly ever worries about Gödel"
 - Newell-Simon-Shaw [1956] Logical Theory Machine: subset of Principia
 - Type Theory [1970s] Martin-Löf, Coquand, ... Coq: 4-color (2005);
 Feit-Thompson Thm (2012); Lean (2013); Univalent Foundation / Homotopy
 Type Theory [2006-] Voevodsky

Buzzard: "Future of Maths" 2019, ICM 2022 Davenport: ICM 2018

"Computer Assisted Proofs" Szegedy: more extreme view, computers >

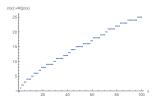
humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

We can call this Bottom-up Mathematics

イロト イポト イヨト イヨト

How does one *DO* mathematics, II ?

- Historically, Maths perhaps more Top-Down: practice before foundation
 - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra . . .
 - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

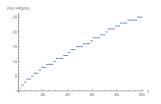


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

 BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, II ?

- Historically, Maths perhaps more Top-Down: practice before foundation
 - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
 - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

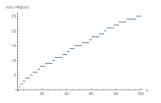


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

 BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, II ?

- Historically, Maths perhaps more Top-Down: practice before foundation
 - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
 - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16



(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

• BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

• NOISELESS Data: different from real-world data to which ML is usually applied; If I gave you 100,000 cases of

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?

• NOISELESS Data: different from real-world data to which ML is usually applied; If I gave you 100,000 cases of

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?

• NOISELESS Data: different from real-world data to which ML is usually applied; If I gave you 100,000 cases of

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?

my fantastic students Jiakang Bao, Elli Heyes, Ed Hirst Tejas Acharya, Daatta Aggrawal, Malik Amir,

Kieran Bull, Lucille Calmon, Siqi Chen, Suvajit Majumder, Maks Manko, Toby Peterken,

Juan Pérez-Ipiña, Max Sharnoff, Yan Xiao my wonderful collaborators

Physics: Guillermo Arias-Tamargo, David Berman, Heng-Yu Chen, Andrei Constantin, Sebastián Franco, Vishnu Jejjala,

Seung-Joo Lee, Andre Lukas, Shailesh Lal, Brent Nelson, Diego Rodriguez-Gomez, Zaid Zaz

Algebraic Geometry: Anthony Ashmore, Challenger Mishra, Rehan Deen, Burt Ovrut

Number Theory: Laura Alessandretti, Andrea Baronchelli, Kyu-Hwan Lee, Tom Oliver, Alexey Pozdnyakov, Drew Sutherland,

Eldar Sultanow

Representation Theory: Mandy Cheung, Pierre Dechant, Minhyong Kim, Jianrong Li, Gregg Musiker

Combinatorics: Johannes Hofscheier, Alexander Kasprzyk, Shiing-Tung Yau

< □ > < 同 > < 回 > < 回 >

Clearly useful for maths and physics

- conjecture formulation: e.g.
 - '19 YHH-Kim: separating hyperplane simple/non-simple groups; open
 - '19 Brodie-Constantin-Lukas: exact formulae for cohomo surf.; proved.
 - '20 YHH-Lee-Oliver: L-coefs and integer pt./torsion on ell; Known.
 - '20 Craven-Jejjala-Par: Jones poly best-fit function; open
 - '22 DeepMind-Oxford-Sydney, Nature: Volume bounds for knots; proved
- speed-ups & accuracies: e.g.,
 - computing/estimating (top.inv., charges, etc) MUCH FASTER
 - '19 Ashmore-YHH-Ovrut: speed up Donaldson alg@CY metric 10-100
 - '20 Douglas et al., Anderson et al. improves Donaldson 10-100 times

A D F A A F F A

Please submit

Launching in 2023		
IJDSMS	INTERNATIONAL JOURNAL OF DATA SCIENCE IN THE MATHEMATICAL	
Calling for Papers	SCIENCES	
	Prof. Yang-Hui He LINS, Royal Institution & Merton College, Dutord	
Editor-in-Chief Yang-Hui He		
London Institute for Mathematical Sciences & Merton College, University of Oxford email: hey@maths.ox.ac.uk		
	World Scientific	
More Information:		

メロト スピト メヨト メヨト

The London Institute for Mathematical Sciences

- UK's only independent research institute for maths; modelled after IAS, Princeton
- Founded in 2011 by Dr. Thomas Fink
- Housed in the Faraday Suites of the Royal Institution of Great Britain
- 1 of 23 themes: AI for Maths Discovery

https://lims.ac.uk/event/ai-assisted-maths-discovery/

Just established:

Arnold Felowships Landau Fellowships





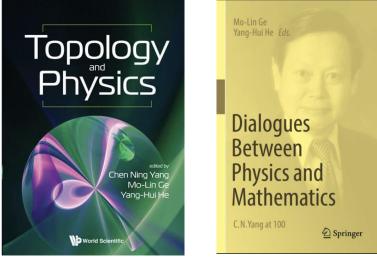




Obrigada!



・ロト ・回ト ・ヨト



CN Yang, ML Ge & YH He, ed, World Scien-

tific, 2019 contributions: Atiyah, Dijkgraaf,

Kim, Penrose, Witten, et al.

ML Ge & YH He, ed, Springer-Nature, 2022

contributions: Drinfeld, Leggett, Manin,

Penrose, Polyakov, Wilczek, Witten, et al.

Just came out in 2020 ...



CalabiYau the Game



< □ > < □ > < □ > < □ > < □ >



Back to Serious Geometry

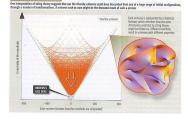
游戏介绍(from https://www.9k9k.com/shouyou/klbq/) 《卡拉比丘》是一款宏大世界观的动作游戏。采用5V5的战斗模式, 玩家需要选择自己的阵营,操控英雄探索地图,与队友密切协作, 战胜敌人玩家即可获得比赛的胜利,更有上百位美少女英雄等待你的召 唤!

Triadophilia

Exact (MS)SM Particle Content from String Compactification

- [Braun-YHH-Ovrut-Pantev, Bouchard-Cvetic-Donagi 2005] first exact MSSM
- [Anderson-Gray-YHH-Lukas, 2007-] use alg./comp. algebraic geo & sift
- Anderson-Gray-Lukas-Ovrut-Palti ~ 200 in 10¹⁰ MSSM Stable Sum of Line Bundles over CICYs (Oxford-Penn-Virginia 2012-)

Constantin-YHH-Lukas '19: 10^{23} exact MSSMs (by extrapolation on above set)?



_ /. /.....

5/1/2008 feature]

Candelas-de la Ossa-YHH-Szendroi

A Special Corner [New Scientist,

"Triadophilia: A Special Corner of the Land-

• • • • • • • • • • • • •

scape" ATMP, 2008

Back to String Landscape

Computing Geometrical Invariants

• Recall Hodge decomposition $H^{p,q}(X) \simeq H^q(X, \wedge^p T^\star X) \rightsquigarrow$

 $H^{1,1}(X) = H^1(X, T_X^*), \qquad H^{2,1}(X) \simeq H^{1,2} = H^2(X, T_X^*) \simeq H^1(X, T_X)$

• Euler Sequence for subvariety *X* ⊂ *A* is short exact:

$$0 \to T_X \to T_M|_X \to N_X \to 0$$

• Induces long exact sequence in cohomology :

• Need to compute Rk(d), cohomology and $H^i(X, T_A|_X)$ (Cf. Hübsch)

YANG-HUI HE (London/Oxford/Nankai)

イロト イポト イヨト イヨト