Universes as Bigdata,

from String Theory to Modern Geometry to Machine-Learning

YANG-HUI HE

Dept of Mathematics, City, **University of London**

Merton College, **University of Oxford**

School of Physics, **NanKai University**

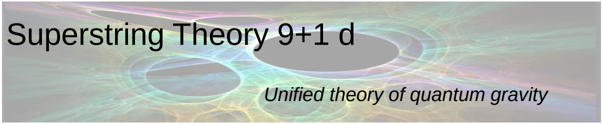
International Congress of Mathematical Software 2020

Braunschweig July, 2020

The Geometrization Programme

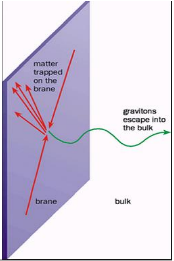
- Algebraic/differential geometry/topology : the right language for physics
  - Gravity $\sim$ Ricci 2-form of the Tangent bundles;
  - Elementary Particles $\sim$ irreducible representations of the Lorentz group and sections of bundles with Lie structure group;
  - Interactions $\sim$ Tensor products of sections ...
  - **String theory: brain-child of gauge-gravity geometrization tradition**
- A new exciting era for synergy with (pure & computational) geometry, group theory, combinatorics, number theory: *Sage, M2, GAP, LMFDB, GrDB* are becoming indispensible tools for physicists
- Interdisciplinary enterprise: cross-fertilisation of particle/string theory, phenomenology, pure mathematics, computer algorithms, data-bases, ...

$10 = 4 + 3 \times 2$



I. 6 Large Dim

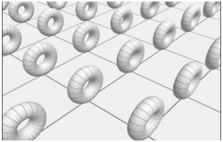
AdS/CFT
Brane World

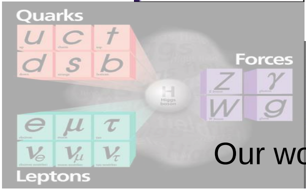


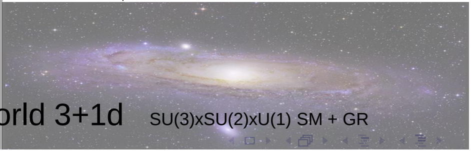
1. *Reduce Dim: 10 = 6+4*
2. *Break SUSY*

II. 6 small dim

Compactification







Our world 3+1d

SU(3)xSU(2)xU(1) SM + GR

Triadophilia: 1984/5

- String Phenomenology [Candelas-Horowitz-Strominger-Witten]: 1985
  - Heterotic string [Gross-Harvey-Martinec-Rohm]: $E_8 \times E_8$ or $SO(32)$, 1984 - 6
  - E_8 accommodates Standard Model of particle physics

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

  - 6 extra dimensions is some 6-dimensional manifold X
    1. not just a real 6-manifold but a complex 3-fold X
    2. X is furthermore Kähler ($g_{\alpha\bar{\beta}} = \partial_\alpha \bar{\partial}_{\bar{\beta}} K$)
    3. X is Ricci flat (vacuum Einstein equations)
    4. Rmk: **there are other classes of solutions** (more later...) but 1,2,3 simplest
- What are such manifolds? *Just so happens that Yau and Strominger were neighbours at IAS in 1985*

Calabi-Yau

- Generalize the trichotomy for complex dim 1 (Riemann surfaces):

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

- HARD for $\dim_{\mathbb{C}} > 1$, luckily, for our class of **Kähler** complex manifolds:
- CONJECTURE [E. Calabi, 1954, 1957]: M compact Kähler (g, ω) and $([R] = [c_1(M)])_{H^{1,1}(M)}$.$\exists!(\tilde{g}, \tilde{\omega})$ s.t. $([\omega] = [\tilde{\omega}])_{H^2(M;\mathbb{R})}$ & $Ricci(\tilde{\omega}) = R$.

Rmk: $c_1(M) = 0 \Leftrightarrow$ Ricci-flat (rmk: Ricci-flat familiar in GR long before strings)

- THEOREM [S-T Yau, 1977-8; Fields 1982] Existence Proof
- Calabi-Yau: Kähler and Ricci-flat (term coined by physicists)

Calabi-Yau

- Generalize the trichotomy for complex dim 1 (Riemann surfaces):

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

- HARD for $\dim_{\mathbb{C}} > 1$, luckily, for our class of **Kähler** complex manifolds:
- CONJECTURE [E. Calabi, 1954, 1957]: M compact Kähler (g, ω) and $([R] = [c_1(M)])_{H^{1,1}(M)}$.$\exists!(\tilde{g}, \tilde{\omega})$ s.t. $([\omega] = [\tilde{\omega}])_{H^2(M;\mathbb{R})}$ & $Ricci(\tilde{\omega}) = R$.

Rmk: $c_1(M) = 0 \Leftrightarrow$ Ricci-flat (rmk: Ricci-flat familiar in GR long before strings)

- THEOREM [S-T Yau, 1977-8; Fields 1982] Existence Proof
- Calabi-Yau: Kähler and Ricci-flat (term coined by physicists)

Calabi-Yau

- Generalize the trichotomy for complex dim 1 (Riemann surfaces):

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

- HARD for $\dim_{\mathbb{C}} > 1$, luckily, for our class of **Kähler** complex manifolds:
- CONJECTURE [E. Calabi, 1954, 1957]: M compact Kähler (g, ω) and $([R] = [c_1(M)])_{H^{1,1}(M)}$.$\exists!(\tilde{g}, \tilde{\omega})$ s.t. $([\omega] = [\tilde{\omega}])_{H^2(M;\mathbb{R})}$ & $Ricci(\tilde{\omega}) = R$.

Rmk: $c_1(M) = 0 \Leftrightarrow$ Ricci-flat (rmk: Ricci-flat familiar in GR long before strings)

- THEOREM [S-T Yau, 1977-8; Fields 1982] Existence Proof
- Calabi-Yau: Kähler and Ricci-flat (term coined by physicists)

Calabi-Yau Manifolds as Algebraic Varieties

- THM: Homog deg $n+1$ in $\mathbb{P}^n$, is Calabi-Yau $\dim_{\mathbb{C}} = n-1$ (adjunction)
- dim$_{\mathbb{C}} = 1$: T^2 as cubic (elliptic curve) in $\mathbb{P}^2$;

  dim$_{\mathbb{C}} = 2$: K3 surface as quartic in $\mathbb{P}^3$
- CY3, immediately get 5 (cyclics): Degree 5 in $\mathbb{P}^4$ (The Quintic Q), [3,3] in $\mathbb{P}^5$, [2,4] in $\mathbb{P}^5$, [2,2,3] in $\mathbb{P}^6$, [2,2,2,2] in $\mathbb{P}^7$
- First physics challenge to algebraic geometry:
  - Particle Spectrum: Generation : $n_{27} = h^1(X, TX) = h^{2,1}_{\bar\partial}(X)$;

    Anti-Generation : $n_{\overline{27}} = h^1(X, TX^*) = h^{1,1}_{\bar\partial}(X)$
  - $\#$ generations of particles $= \chi = 2(h^{1,1} - h^{2,1})$; 1986 Question: Are there Calabi-Yau threefolds with Euler number ± 6? (None of our 5 obvious ones)

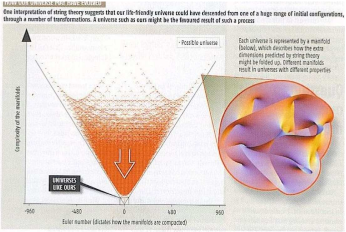
The First Data-sets in Mathematical Physics/Geometry

- [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, $\chi \in [-960, 960]$

Triadophilia: recent progress

Exact Standard Model 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 10^{10} MSSM Stable Sum of Line Bundles over CICYs (Oxford-Penn-Virginia 2012-)



A Special Corner?

[New Scientist, 5/1/2008 feature]

Candelas-de la Ossa-YHH-Szendroi

"Triadophilia: A Special Corner of the Landscape" ATMP, 2008

Where we stand ...

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** ($10^{9\sim10}$ entries typically) links

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**

A Wild Question

- Typical Problem in String Theory/Algebraic Geometry:

INPUT		*OUTPUT*
integer tensor	$\longrightarrow$	integer

- 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, *Phys Lett B 774, 2017*

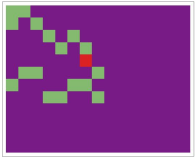
  Science feature article, Aug, vol 365 issue 6452 :

  Experimentally, it seems to be the case for many situations in geometry and beyond. (cf. YHH *CY Landscape: from Geometry, to Physics, to ML* 1812.02893, Springer Textbook, to appear)

NN Doesn't Care/Know about Algebraic Geometry

- Hodge Number of a Complete Intersection CY is the association rule, e.g.

$$X = \begin{pmatrix} 1&1&0&0&0&0&0&0 \\ 1&0&1&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0 \\ 0&0&0&0&1&0&1&0 \\ 0&0&0&0&0&0&2&0 \\ 0&1&1&0&0&0&0&1 \\ 1&0&0&0&0&1&1&0 \\ 0&0&0&1&1&0&0&1 \end{pmatrix}, \qquad h^{1,1}(X) = 8 \quad \rightsquigarrow$$



$\longrightarrow 8$

CICY is 12×15 integer matrix with entries $\in [0,5]$ is simply represented as a 12×15 pixel image of 6 colours (proper way: sequence chasing) Proper Way

- Cross-Validation:
  - Take samples of $X \to h^{1,1}$
  - train a NN, or SVM
  - Validation on *unseen* $X \to h^{1,1}$

Experiments ((precision, confidence) @ 80-20 cross-val)

- YHH (1706.02714) Bull-YHH-Jejjala-Mishra (1806.03121, 1903.03113): Hodge numbers of CICYs, $(0.85, 0.9)$;
- YHH-SJ Lee (1904.08530): Distinguish Elliptic Fibrations $(0.99, 0.99)$;
- Ashmore-YHH-Ovrut (1910.08605): ML Donaldson algorithm for (balanced) CY metric $\sim 10^2$ speedup;
- YHH-MY Kim (1905.02263): Use GAP database, recognizing simple groups: $(0.96, 0.95)$ distinguishing Cayley tables from Latin squares $(0.9, 0.9)$
- Bao-Franco-YHH-Hirst-Musiker-Xiao (2006.10783): Recognize cluster mutations $(0.9, 0, 8) \to (1.0, 1.0)$
- YHH-ST Yau (2006.16619): Finite graphs, recognizing acyclic, Ricci-flat, planarity, etc: $(0.8, 0.7) \to (0.95, 0.91)$ Reprobates

Summary and Outlook

PHYSICS
- The string landscape now solidly resides in the age of Big Data
- Use Machine-Learning as Classifier & Predictor

MATHS
- somewhat bypassing the expensive steps of long sequence-chasing, Gröbner bases, dual cones/combinatorics and getting the right answer @ high rate (only probabilistically doing NP-Hard); 100% $\rightsquigarrow$ new conjectures/formulae/algorithms;
- Hierarchy of Difficulty ML struggles with:
  numerical $<$ **algebraic geometry over** $\mathbb{C}$ $<$ **combinatorics/algebra** $<$ **number theory**

Various Databases

- CICYs: resurrected Anderson-Gray-YHH-Lukas, http://www-thphys.physics.ox.ac.uk/projects/CalabiYau/cicylist/index.html
- Kreuzer-Skarke: http://hep.itp.tuwien.ac.at/~kreuzer/CY/
  - new PALP: Braun-Walliser: ArXiv 1106.4529
  - Triangulations: Altmann-YHH-Jejjala-Nelson: http://www.rossealtman.com/
- cf. Graded Rings: Brown, Kasprzyk, et al. http://www.grdb.co.uk/

Return

Computing Hodge Numbers: Sketch

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

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

- Euler Sequence for subvariety $X \subset A$ is short exact:

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

- Induces long exact sequence in cohomology:

$$\begin{array}{ccccccccc}
0 & \to & H^0(X, T_X) \overset{0}{\nearrow} & \to & H^0(X, T_A|_X) & \to & H^0(X, N_X) & \to \\
 & \to & \boxed{H^1(X, T_X)} & \stackrel{d}{\to} & H^1(X, T_A|_X) & \to & H^1(X, N_X) & \to \\
 & \to & H^2(X, T_X) & \to & \ldots & & &
\end{array}$$

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

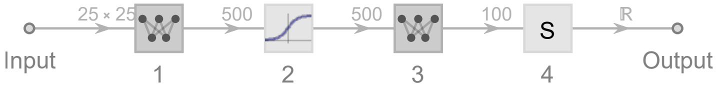
Back to ML

Reprobates

Expect anything involving properties of primes to do badly:

- YHH (1706.02714): predicting primes $(0.5, 0.0)$ (completely random) . . .
- Alessandretti-Baronchelli-YHH (1911.02008) BSD using LMFdb: rank from Weierstrass coef, random; correlation @ BSD quantities, marginally better

at least for standard classifers (boosted trees, SVMs, ...) and simple NN which does algebraic geometry so well:



Return