Overview of the String Landscape

Michael R. Douglas

[1] Simons Center / Stony Brook University

Northeastern, November 30, 2017

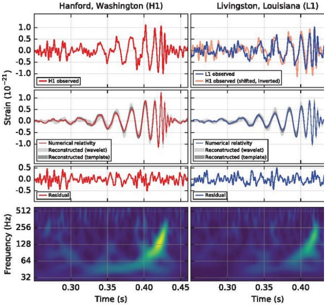
Abstract

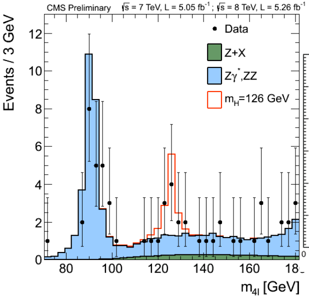
We introduce string phenomenology and the string landscape for a general scientific audience, and make the case that the study of the string landscape is a multidisciplinary field which should involve computer scientists, data scientists and computational mathematicians.

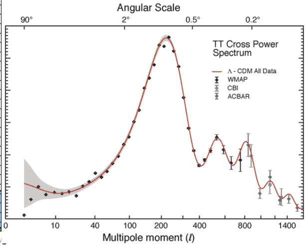
Outline

1. The state of fundamental physics
2. What is the string landscape?
3. Interactions between theoretical physics and computer science
4. Computational mathematical science as a collaborative effort

The state of fundamental physics







We have a theory that describes all experiments to date: general relativity (GR), the Standard Model (SM) with neutrino masses, inflation plus dark matter and energy.

This cannot be a complete theory as GR is not quantum, and the inflation/dark matter/dark energy part is not precise.
Energy is inverse distance, so substructure is only visible at higher energies. What do we know about physics at higher energies?

Important energy scales:

- Planck scale: 10^{19} GeV
- scale of observed inflation: between 10^5 GeV and 10^{15} GeV
- electroweak breaking: 100 GeV
- QCD scale: 0.1 GeV

Well motivated but speculative important scales:

- GUT scale: 10^{16} GeV
- Scale of supersymmetry breaking: greater than 10^4 GeV

The most important things we know about physics at higher energies are **consistency conditions**. It is not easy to satisfy the rules of quantum mechanics and have locality of interactions in a precise mathematical framework. While there are many ways to do it, the ways can be systematically classified using **effective field theory** (EFT).

This cannot be a complete theory as GR is not quantum, and the inflation/dark matter/dark energy part is not precise.
Energy is inverse distance, so substructure is only visible at higher energies. What do we know about physics at higher energies?

Important energy scales:

- Planck scale: 10^{19} GeV
- scale of observed inflation: between 10^5 GeV and 10^{15} GeV
- electroweak breaking: 100 GeV
- QCD scale: 0.1 GeV

Well motivated but speculative important scales:

- GUT scale: 10^{16} GeV
- Scale of supersymmetry breaking: greater than 10^4 GeV

The most important things we know about physics at higher energies are **consistency conditions**. It is not easy to satisfy the rules of quantum mechanics and have locality of interactions in a precise mathematical framework. While there are many ways to do it, the ways can be systematically classified using **effective field theory** (EFT).

This cannot be a complete theory as GR is not quantum, and the inflation/dark matter/dark energy part is not precise.
Energy is inverse distance, so substructure is only visible at higher energies. What do we know about physics at higher energies?

Important energy scales:

- Planck scale: 10^{19} GeV
- scale of observed inflation: between 10^5 GeV and 10^{15} GeV
- electroweak breaking: 100 GeV
- QCD scale: 0.1 GeV

Well motivated but speculative important scales:

- GUT scale: 10^{16} GeV
- Scale of supersymmetry breaking: greater than 10^4 GeV

The most important things we know about physics at higher energies are **consistency conditions**. It is not easy to satisfy the rules of quantum mechanics and have locality of interactions in a precise mathematical framework. While there are many ways to do it, the ways can be systematically classified using **effective field theory** (EFT).

This cannot be a complete theory as GR is not quantum, and the inflation/dark matter/dark energy part is not precise.
Energy is inverse distance, so substructure is only visible at higher energies. What do we know about physics at higher energies?

Important energy scales:

- Planck scale: 10^{19} GeV
- scale of observed inflation: between 10^5 GeV and 10^{15} GeV
- electroweak breaking: 100 GeV
- QCD scale: 0.1 GeV

Well motivated but speculative important scales:

- GUT scale: 10^{16} GeV
- Scale of supersymmetry breaking: greater than 10^4 GeV

The most important things we know about physics at higher energies are **consistency conditions**. It is not easy to satisfy the rules of quantum mechanics and have locality of interactions in a precise mathematical framework. While there are many ways to do it, the ways can be systematically classified using **effective field theory** (EFT).

This cannot be a complete theory as GR is not quantum, and the inflation/dark matter/dark energy part is not precise.
Energy is inverse distance, so substructure is only visible at higher energies. What do we know about physics at higher energies?

Important energy scales:

- Planck scale: 10^{19} GeV
- scale of observed inflation: between 10^{5} GeV and 10^{15} GeV
- electroweak breaking: 100 GeV
- QCD scale: 0.1 GeV

Well motivated but speculative important scales:

- GUT scale: 10^{16} GeV
- Scale of supersymmetry breaking: greater than 10^{4} GeV

The most important things we know about physics at higher energies are **consistency conditions**. It is not easy to satisfy the rules of quantum mechanics and have locality of interactions in a precise mathematical framework. While there are many ways to do it, the ways can be systematically classified using **effective field theory** (EFT).

Effective field theory

We can get most of the physical predictions of a consistent fundamental theory by first deriving its EFT. This describes the particles and interactions which can be seen by doing experiments up to a "cutoff" energy E_{max}. The Standard Model is an EFT valid at energies up to $\mathrm{E_{max}} \sim 1\ \mathrm{TeV}$ and possibly much higher. But we do not know whether it is complete. If we had the complete EFT for $\mathrm{E_{max}} = 13\ \mathrm{TeV}$, we would know what particles to look for (at LHC, in dark matter searches, neutrino experiments etc.) and tests to make.

An EFT is specified by a gauge group G, two group representations $\mathrm{R_f}$ and $\mathrm{R_b}$, and a Lagrangian. Generally, each factor of the group representation describes a particle. The Lagrangian depends on a finite set of numbers, the particle masses and couplings. There are some manifest symmetries (*e.g.* rotations between fields of the same type) and some highly nonobvious symmetries (dualities).

Effective field theory

We can get most of the physical predictions of a consistent fundamental theory by first deriving its EFT. This describes the particles and interactions which can be seen by doing experiments up to a “cutoff” energy E_{max}. The Standard Model is an EFT valid at energies up to $E_{max} \sim 1\ \mathrm{TeV}$ and possibly much higher. But we do not know whether it is complete. If we had the complete EFT for $E_{max} = 13\ \mathrm{TeV}$, we would know what particles to look for (at LHC, in dark matter searches, neutrino experiments etc.) and tests to make.

An EFT is specified by a gauge group G, two group representations R_f and R_b, and a Lagrangian. Generally, each factor of the group representation describes a particle. The Lagrangian depends on a finite set of numbers, the particle masses and couplings. There are some manifest symmetries (*e.g.* rotations between fields of the same type) and some highly nonobvious symmetries (dualities).

For now, the main point is that we can specify a candidate physical theory using a relatively small data structure: of order N^k numbers with $k = 2, 3, 4, 5$ for an EFT with N particles.

Using the renormalization group, the parameters of an EFT can be divided into “relevant,” “marginal” and “irrelevant.” The terminology comes from statistical physics and needs to be reinterpreted in fundamental physics:

- marginal $=$ scale invariant (up to logarithms)
- relevant $=$ unnatural (needs explanation in fundamental theory)
- irrelevant $=$ signal of physics at $E > E_{max}$, perhaps new particles

Bare mass terms are relevant, so we prefer theories in which particles get their mass “dynamically” through interactions with other fields.

Irrelevant couplings are classified by dimension D and energy scale Λ. If we can do experiments at an energy E, then such an operator will generally have effects of size $(E/\Lambda)^{D-4}$.

For now, the main point is that we can specify a candidate physical theory using a relatively small data structure: of order N^k numbers with $k = 2, 3, 4, 5$ for an EFT with N particles.

Using the renormalization group, the parameters of an EFT can be divided into "relevant," "marginal" and "irrelevant." The terminology comes from statistical physics and needs to be reinterpreted in fundamental physics:

- marginal $=$ scale invariant (up to logarithms)
- relevant $=$ unnatural (needs explanation in fundamental theory)
- irrelevant $=$ signal of physics at $E > E_{max}$, perhaps new particles

Bare mass terms are relevant, so we prefer theories in which particles get their mass "dynamically" through interactions with other fields.

Irrelevant couplings are classified by dimension D and energy scale Λ. If we can do experiments at an energy E, then such an operator will generally have effects of size $(E/\Lambda)^{D-4}$.

For now, the main point is that we can specify a candidate physical theory using a relatively small data structure: of order N^k numbers with $k = 2, 3, 4, 5$ for an EFT with N particles.

Using the renormalization group, the parameters of an EFT can be divided into "relevant," "marginal" and "irrelevant." The terminology comes from statistical physics and needs to be reinterpreted in fundamental physics:

- marginal $=$ scale invariant (up to logarithms)
- relevant $=$ unnatural (needs explanation in fundamental theory)
- irrelevant $=$ signal of physics at $E > E_{max}$, perhaps new particles

Bare mass terms are relevant, so we prefer theories in which particles get their mass "dynamically" through interactions with other fields.

Irrelevant couplings are classified by dimension D and energy scale Λ. If we can do experiments at an energy E, then such an operator will generally have effects of size $(E/\Lambda)^{D-4}$.

The SM has gauge group $SU(3) \times SU(2) \times U(1)$. The quarks and leptons are fermions

$$R_f = \oplus_{g=1}^{3} \left((3,2)_{1/6} \oplus (\bar{3},1)_{-2/3} \oplus (\bar{3},1)_{1/3} \oplus (1,2)_{1/2} \oplus (1,1)_{-1} \right)$$

while the Higgs boson has $R_b = (1,2)_{1/2}$. There is 1 relevant parameter μ and 18 marginal parameters $g_i, y_{ab}, y'_{ab}, \lambda$ (after using rotations $\psi \to U\psi$ and leaving out neutrino masses).

$$\begin{aligned} \mathcal{L} &= \sum_{i=1,2,3} \frac{1}{g_i^2} \mathrm{tr}\, F_i^2 + \sum \bar{\psi}_a \not{D} \psi_a + y_{ab} \bar{\psi}_a \phi \psi_b + y'_{ab} \bar{\psi}_a \phi^* \psi_b \\ &\quad + |D\phi|^2 - \lambda \left(|\phi|^2 - \mu^2 \right)^2 . \end{aligned}$$

There are a few $D = 5$ and many $D = 6$ irrelevant operators which could be detected by precision experiments ($g - 2$, FCNC's, etc.)

The relevant parameter μ was one of the original motivations for the supersymmetric standard model. This has ~ 100 marginal parameters.

The SM has gauge group $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. The quarks and leptons are fermions

$$R_f = \oplus_{g=1}^{3} \left((3,2)_{1/6} \oplus (\bar{3},1)_{-2/3} \oplus (\bar{3},1)_{1/3} \oplus (1,2)_{1/2} \oplus (1,1)_{-1}\right)$$

while the Higgs boson has $R_b = (1,2)_{1/2}$. There is 1 relevant parameter μ and 18 marginal parameters $g_i, y_{ab}, y'_{ab}, \lambda$ (after using rotations $\psi \to U\psi$ and leaving out neutrino masses).

$$\begin{aligned}\mathcal{L} &= \sum_{i=1,2,3} \frac{1}{g_i^2} \mathrm{tr}\, F_i^2 + \sum \bar{\psi}_a \not{D} \psi_a + y_{ab} \bar{\psi}_a \phi \psi_b + y'_{ab} \bar{\psi}_a \phi^* \psi_b \\ &\quad + |D\phi|^2 - \lambda \left(|\phi|^2 - \mu^2\right)^2 .\end{aligned}$$

There are a few $D = 5$ and many $D = 6$ irrelevant operators which could be detected by precision experiments ($g - 2$, FCNC's, etc.)

The relevant parameter μ was one of the original motivations for the supersymmetric standard model. This has ~ 100 marginal parameters.

Many, many ideas for physics beyond the standard model (BSM) physics have been suggested. Besides explaining the relevant parameter μ, other motivations include to explain (or just model) the dark matter or dark energy, to explain neutrino masses, to explain baryogenesis or other features of early cosmology, or just to fit with discoveries which seemed (at the time) to contradict the SM. Almost all ideas which fit current observation lead to an EFT which determines their predictions.

Thus, to summarize the situation, the main way we have to test a candidate fundamental theory of physics is to derive its EFT and from this derive its predictions for new particles or interactions, which can be tested against experiment or observation. The Standard Model is an EFT which has passed many, many tests. But there is a good deal of room for new physics – either at energies $E \gg 1\mathrm{TeV}$ or in new sectors which couple very weakly to the Standard Model particles. The goal of string phenomenology is to identify the new physics which can – or cannot – come from string theory and propose such tests.

Many, many ideas for physics beyond the standard model (BSM) physics have been suggested. Besides explaining the relevant parameter μ, other motivations include to explain (or just model) the dark matter or dark energy, to explain neutrino masses, to explain baryogenesis or other features of early cosmology, or just to fit with discoveries which seemed (at the time) to contradict the SM. Almost all ideas which fit current observation lead to an EFT which determines their predictions.

Thus, to summarize the situation, the main way we have to test a candidate fundamental theory of physics is to derive its EFT and from this derive its predictions for new particles or interactions, which can be tested against experiment or observation. The Standard Model is an EFT which has passed many, many tests. But there is a good deal of room for new physics – either at energies $\mathrm{E} \gg 1\mathrm{TeV}$ or in new sectors which couple very weakly to the Standard Model particles. The goal of string phenomenology is to identify the new physics which can – or cannot – come from string theory and propose such tests.

String/M theory

String/M theory is the first and the best candidate we have for a theory underlying all of fundamental physics:

- It unifies gravity and Yang-Mills theories with matter.
- Thanks to supersymmetry, it does not have the UV divergences of field theoretic quantum gravity in $\mathrm{D} > 2$, while still preserving Lorentz invariance and the concept of continuous spacetime.
- It realizes maximal symmetries and other exceptional structures: maximal supergravity, $\mathrm{N} = 4$ SYM, E_8, ...
- It realizes a surprising network of dualities which unify many ideas in theoretical physics: strong$\leftrightarrow$weak, gauge$\leftrightarrow$gravity, ...
- Although it is naturally formulated in 10 and 11 space-time dimensions, it is not hard to find solutions which are a direct product of 4d space-time with a small compact space (the “extra dimensions”), and for which the effective 4d physics at low energies is the Standard Model coupled to gravity.

As we will discuss, string/M theory has a large number of solutions for the shape and size of the extra dimensions. Only a minority of these solutions lead to the Standard Model field content, and those which do lead to a range of values for the cosmological constant, the particle masses and the other fundamental constants.

This enables the anthropic solution to the cosmological constant problem. Anthropic ideas can help answer other questions about “why is the universe suited for our existence?” Many constraints on the fundamental laws (*e.g.* for stability of matter and of large scale structure) do not seem to follow from top-down considerations, and anthropic explanations are arguably more convincing.

It also makes it more difficult to get definite predictions from the theory. To test the theory we want to make predictions for physics beyond the Standard Model. Now there are many negative predictions: possible physics which cannot come out of string theory. For example, a time-varying fine structure constant (Banks *et al* 2000). But to make positive predictions, we must know more about the set of solutions.

Quasi-realistic solutions of string/M theory

The study of string compactification began in 1985 with the work of Candelas, Horowitz, Strominger and Witten on compactification of heterotic string theory on a six dimensional Calabi-Yau manifold. They gave convincing examples of compactifications that lead to a four-dimensional supersymmetric grand unified theory with quarks and leptons. Soon three generation models were found.

For many years, the outstanding problems were to eliminate unobserved scalar fields (moduli) and explain the smallness of the cosmological constant. Solutions were proposed in the early 2000's, in particular the Bousso-Polchinski model and the KKLT discussion of moduli stabilization. In particular KKLT pointed out that vacua with positive cosmological constant can at best be metastable – though with extremely long lifetimes. This explains why they do not come out of simple constructions.

Quasi-realistic solutions of string/M theory

The study of string compactification began in 1985 with the work of Candelas, Horowitz, Strominger and Witten on compactification of heterotic string theory on a six dimensional Calabi-Yau manifold. They gave convincing examples of compactifications that lead to a four-dimensional supersymmetric grand unified theory with quarks and leptons. Soon three generation models were found.

For many years, the outstanding problems were to eliminate unobserved scalar fields (moduli) and explain the smallness of the cosmological constant. Solutions were proposed in the early 2000's, in particular the Bousso-Polchinski model and the KKLT discussion of moduli stabilization. In particular KKLT pointed out that vacua with positive cosmological constant can at best be metastable – though with extremely long lifetimes. This explains why they do not come out of simple constructions.

Over the years, many constructions of vacua whose EFT is the SM – at least up to the accuracy we can compute it – have been proposed. There fall into several classes:

- IIb, IIa or heterotic strings on CY_3
- M theory on a 7d manifold of G_2 holonomy
- F theory on an elliptically fibered CY_4

Each gives a subset of the landscape of $\mathcal{N} = 1$ supersymmetric vacua with low energy susy breaking. At present F theory gives the largest subset and many speakers at this conference will use it.

As we will briefly outline, and as the speakers at this conference will explain in detail, within each construction, there are many discrete choices which can be varied: the topology of the extra dimensions; embeddings of branes and other objects; and generalized magnetic flux which determines the shape of the extra dimensions. Each choice leads to a different EFT in four dimensions.

Over the years, many constructions of vacua whose EFT is the SM – at least up to the accuracy we can compute it – have been proposed. There fall into several classes:

- IIb, IIa or heterotic strings on CY_3
- M theory on a 7d manifold of G_2 holonomy
- F theory on an elliptically fibered CY_4

Each gives a subset of the landscape of $\mathcal{N} = 1$ supersymmetric vacua with low energy susy breaking. At present F theory gives the largest subset and many speakers at this conference will use it.

As we will briefly outline, and as the speakers at this conference will explain in detail, within each construction, there are many discrete choices which can be varied: the topology of the extra dimensions; embeddings of branes and other objects; and generalized magnetic flux which determines the shape of the extra dimensions. Each choice leads to a different EFT in four dimensions.

An illustrative example: type IIb strings on a Calabi-Yau threefold with branes. Here are the steps to choose a particular compactification and compute the EFT:

- A CY_3 denoted M. For example, a toric hypersurface (more below).
- Choose D3,5,7,9-branes wrapping holomorphic cycles on M with vector bundles and satisfying tadpole constraint.
- Flux, a choice of complex class in $\mathrm{H}^3(\mathrm{M}, \mathbb{Z} \oplus i\mathbb{Z})$.
- Compute the effective potential, including effects of supersymmetry breaking.
- Choose a local minimum and test it for metastability.
- Compute matter spectrum and couplings.

The result is an EFT. The topological choices (M, brane numbers and cycles) determine the gauge group and matter representations. The flux chooses among a list of similar potentials, and the values of the moduli at the minimum translate into masses and couplings.

An illustrative example: type IIb strings on a Calabi-Yau threefold with branes. Here are the steps to choose a particular compactification and compute the EFT:

- A CY_3 denoted M. For example, a toric hypersurface (more below).
- Choose D3,5,7,9-branes wrapping holomorphic cycles on M with vector bundles and satisfying tadpole constraint.
- Flux, a choice of complex class in $\mathrm{H}^3(\mathrm{M}, \mathbb{Z} \oplus \mathrm{i}\mathbb{Z})$.
- Compute the effective potential, including effects of supersymmetry breaking.
- Choose a local minimum and test it for metastability.
- Compute matter spectrum and couplings.

The result is an EFT. The topological choices (M, brane numbers and cycles) determine the gauge group and matter representations. The flux chooses among a list of similar potentials, and the values of the moduli at the minimum translate into masses and couplings.

The different vacua make many different predictions for physics beyond the SM, varying along many axes:

- The radius of M or Kaluza-Klein scale R_{KK} is the distance below which gravity no longer satisfies an inverse square law.
- All known families of metastable compactifications are supersymmetric at high energy, but the breaking scale $\mathrm{M}_{\mathrm{susy}}$ can vary widely. The number distribution is probably $\sim \mathrm{dM}_{\mathrm{susy}}/\mathrm{M}_{\mathrm{susy}}$.
- There is a "topological complexity" axis having to do with numbers of homology cycles, distinct branes, and so on: call this number b. This translates into numbers of gauge groups and matter sectors (most of which can be hidden) in the low energy field theory. This number distribution is probably $\sim \mathrm{C}^{\mathrm{b}}$ for some $\mathrm{C} \sim 10^2$–10^4.
- Idiosyncratic properties of string theory. For example, F theory and heterotic string theory seem to favor GUTs, while certain intersecting brane models seem to favor three generations of matter.

These general properties can lead to observable predictions:

- The KK and susy breaking scales lead directly to predictions for masses of new particles.
- Similarly, GUTs predict proton decay and certain neutrino mass structures; heavy generations might be observable; some vacua have "leptoquarks" or other heavy exotic matter; etc.
- "Complex" compactifications will have many matter sectors and usually many axions (very light particles which gain their mass from nonperturbative effects). These must be hidden to escape existing bounds.

Nevertheless hidden sectors can have observable effects. One of the most interesting is superradiance of rotating black holes (Arvanitaki *et al* `0905.4720`) which may rule out axions in certain mass windows.

These general properties can lead to observable predictions:

- The KK and susy breaking scales lead directly to predictions for masses of new particles.
- Similarly, GUTs predict proton decay and certain neutrino mass structures; heavy generations might be observable; some vacua have "leptoquarks" or other heavy exotic matter; etc.
- "Complex" compactifications will have many matter sectors and usually many axions (very light particles which gain their mass from nonperturbative effects). These must be hidden to escape existing bounds.

Nevertheless hidden sectors can have observable effects. One of the most interesting is superradiance of rotating black holes (Arvanitaki *et al* `0905.4720`) which may rule out axions in certain mass windows.

Vacuum selection

With all of these possibilities, the most important question is, what determines which of these vacua and thus which EFT describes our universe ? The short answer is,

We don't know!

But if there are principles which select or prefer certain vacua, almost certainly, they will come from the study of early cosmology. There is good evidence that our universe underwent a period of inflation, and not all string vacua can be produced this way. Inflation can also lead to predictions for the density of primordial fluctuations (in our universe 10^{-5}) and for structure formation, which might rule out candidates.

In addition, someday we may understand the initial conditions and the dynamics which creates the multiverse, which could lead to a "measure factor" or probability distribution over vacua. We will say much more about this in our second talk at this workshop.

Vacuum selection

With all of these possibilities, the most important question is, what determines which of these vacua and thus which EFT describes our universe ? The short answer is,

We don't know!

But if there are principles which select or prefer certain vacua, almost certainly, they will come from the study of early cosmology. There is good evidence that our universe underwent a period of inflation, and not all string vacua can be produced this way. Inflation can also lead to predictions for the density of primordial fluctuations (in our universe 10^{-5}) and for structure formation, which might rule out candidates.

In addition, someday we may understand the initial conditions and the dynamics which creates the multiverse, which could lead to a "measure factor" or probability distribution over vacua. We will say much more about this in our second talk at this workshop.

Vacuum selection

With all of these possibilities, the most important question is, what determines which of these vacua and thus which EFT describes our universe ? The short answer is,

We don't know!

But if there are principles which select or prefer certain vacua, almost certainly, they will come from the study of early cosmology. There is good evidence that our universe underwent a period of inflation, and not all string vacua can be produced this way. Inflation can also lead to predictions for the density of primordial fluctuations (in our universe 10^{-5}) and for structure formation, which might rule out candidates.

In addition, someday we may understand the initial conditions and the dynamics which creates the multiverse, which could lead to a "measure factor" or probability distribution over vacua. We will say much more about this in our second talk at this workshop.

Vacuum selection

With all of these possibilities, the most important question is, what determines which of these vacua and thus which EFT describes our universe ? The short answer is,

We don't know!

But if there are principles which select or prefer certain vacua, almost certainly, they will come from the study of early cosmology. There is good evidence that our universe underwent a period of inflation, and not all string vacua can be produced this way. Inflation can also lead to predictions for the density of primordial fluctuations (in our universe 10^{-5}) and for structure formation, which might rule out candidates.

In addition, someday we may understand the initial conditions and the dynamics which creates the multiverse, which could lead to a "measure factor" or probability distribution over vacua. We will say much more about this in our second talk at this workshop.

Some candidate vacuum selection principles which are or may be relevant:

- The anthropic principle. This is a selection effect – the fact that our universe has the properties required for us to exist, does not require further explanation. However the list of “properties required for us to exist” is not totally clear. It seems clear that the cosmological constant must be small to get a large universe, that some form matter must be stable, that matter must be able to make complicated bound states (chemistry) and that some sort of free energy source must spontaneously form (stars). One often simplifies this to the claim that the the four forces of nature and the first generation of matter (electron, up, and down quark) must be as in the SM.

- The anthropic principle.
- The entropic principle. Given two candldate laws of physics A and B which satisfy the anthropic principle, if more vacua lead to A than B, A is favored. This leads to counting of vacua as the important theoretical problem to solve.
- The principle of mediocrity. This states that we live in a "typical" universe. This is different from the entropic principle as we grant that there can be a nontrivial probability distribution on the vacua. Much work has been done to derive this from the theory of eternal inflation.

In any case, drawing predictions from any of these principles requires a good picture of the set of all vacua and all EFTs. This has two parts: a statistical survey of the vacua, and a matrix or connectivity graph of tunneling rates (see my other talk).

- The anthropic principle.
- The entropic principle. Given two candIdate laws of physics A and B which satisfy the anthropic principle, if more vacua lead to A than B, A is favored. This leads to counting of vacua as the important theoretical problem to solve.
- The principle of mediocrity. This states that we live in a "typical" universe. This is different from the entropic principle as we grant that there can be a nontrivial probability distribution on the vacua. Much work has been done to derive this from the theory of eternal inflation.

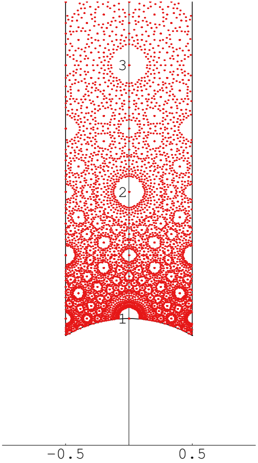
In any case, drawing predictions from any of these principles requires a good picture of the set of all vacua and all EFTs. This has two parts: a statistical survey of the vacua, and a matrix or connectivity graph of tunneling rates (see my other talk).

- The anthropic principle.
- The entropic principle. Given two candldate laws of physics A and B which satisfy the anthropic principle, if more vacua lead to A than B, A is favored. This leads to counting of vacua as the important theoretical problem to solve.
- The principle of mediocrity. This states that we live in a “typical” universe. This is different from the entropic principle as we grant that there can be a nontrivial probability distribution on the vacua. Much work has been done to derive this from the theory of eternal inflation.

In any case, drawing predictions from any of these principles requires a good picture of the set of all vacua and all EFTs. This has two parts: a statistical survey of the vacua, and a matrix or connectivity graph of tunneling rates (see my other talk).

Statistical results

Given a list of EFTs, we can derive its statistics. Here is a set of EFT's all with the same matter content, but in which the complexified gauge coupling $\tau = 4\pi i/g^2 + 2\pi\theta$ takes a discrete set of possible values.



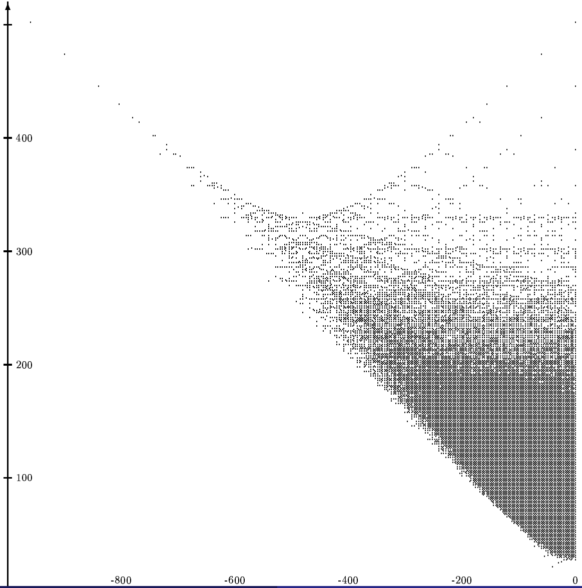
Flux vacua $DW = 0$ with $W = a\tau + b$ and $K = -\mathrm{Im}\ \log\tau$, from Denef and Douglas, 2004.

This graph was obtained by enumerating one solution of $a_1 b_2 - a_2 b_1 = L$ in each $SL(2,\mathbb{Z})$ orbit, taking $\tau = -(b_1 - i b_2)/(a_1 - i a_2)$ and mapping it back to the fundamental region.

The density of vacua is roughly $\pi L d^2\tau/(\mathrm{Im}\ \tau)^2$ as can be found analytically.

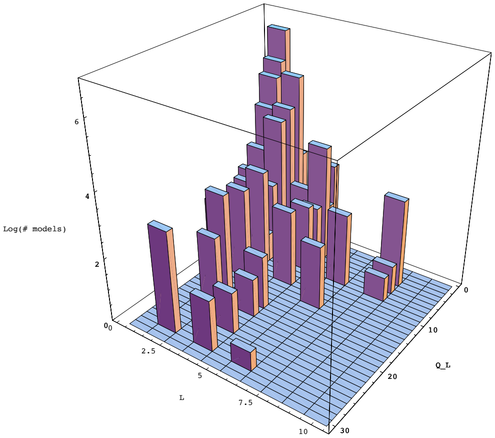
The total number of vacua is $N = 2\sigma(L)$, where $\sigma(L)$ is the sum of the divisors of L. Its large L asymptotics are $N \sim \pi^2 L/6$.

Kreuzer-Skarke survey of toric CY_3's



In this figure, the vertical axis is $b^{1,1} + b^{2,1}$, the number of CY moduli. The horizontal axis is the Euler character $\chi = 2b^{1,1} - 2b^{2,1}$, twice the number of generations in $(2,2)$ heterotic compactification. Each dot represents one or more CY_3's. Mirror symmetry is the reflection symmetry with the (omitted) $\chi > 0$ quadrant.

One can get lists of EFTs with different gauge groups and matter content by working within one of the specific constructions and either enumerating or sampling the choices. Many surveys of this type have been done.



One early result was Gmeiner *et al* `arXiv:hep-th/0510170` which worked with IIa intersecting branes on a toroidal orientifold. They found that, within their (non GUT) ensemble, while the SM gauge group was not hard to realize, three generations of quarks and leptons were very much disfavored (order 10^{-8}) over 2 and 4. However, other brane constructions favor 3.

Mathematical ingredients

- Finite group theory – classification, representation theory, group cohomology.
- Lie algebras and groups – likewise
- Theory of integral lattices.
- Algebraic geometry: construction of manifolds, especially Calabi-Yau manifolds, holomorphic vector bundles and the like.
- Toric geometry – a specific construction which reduces certain problems in algebraic geometry to the combinatorics of polytopes. The Kreuzer-Skarke database of toric hypersurfaces is based on this, as is most work on heterotic string compactification.
- Singularity theory
- Quiver representation theory
- Theory of moduli spaces, arithmetic groups, automorphic functions and forms
- Theory of hypergeometric functions and PDEs.
- Diverse numerical techniques

Summary

To summarize the status of the string landscape as a “top-down” concept which leads to candidate vacua, EFT, and predictions,

- Classification and calculational techniques are moderately well developed but by no means a mature subject.
- In addition people have suggested other relevant classes of 4d vacua (*e.g.* nonsupersymmetric), and in studying cosmology one may need to understand vacua with other numbers of large dimensions. So this is definitely a work in progress.
- On the other hand, the string landscape is a mathematically well-defined structure, someday even rigorously so, and eventually we will understand it.

I believe that it will be largely understood and alternatives to string theory either found or ruled out on the 20-50 year timescale. Thus, although we must always be mindful of the current limits of knowledge, it is reasonable to think that the structures that have been discovered so far will be a significant portion of the final picture.

Summary

To summarize the status of the string landscape as a "top-down" concept which leads to candidate vacua, EFT, and predictions,

- Classification and calculational techniques are moderately well developed but by no means a mature subject.
- In addition people have suggested other relevant classes of 4d vacua (*e.g.* nonsupersymmetric), and in studying cosmology one may need to understand vacua with other numbers of large dimensions. So this is definitely a work in progress.
- On the other hand, the string landscape is a mathematically well-defined structure, someday even rigorously so, and eventually we will understand it.

I believe that it will be largely understood and alternatives to string theory either found or ruled out on the 20-50 year timescale. Thus, although we must always be mindful of the current limits of knowledge, it is reasonable to think that the structures that have been discovered so far will be a significant portion of the final picture.

Summary

To summarize the status of the string landscape as a "top-down" concept which leads to candidate vacua, EFT, and predictions,

- Classification and calculational techniques are moderately well developed but by no means a mature subject.
- In addition people have suggested other relevant classes of 4d vacua (*e.g.* nonsupersymmetric), and in studying cosmology one may need to understand vacua with other numbers of large dimensions. So this is definitely a work in progress.
- On the other hand, the string landscape is a mathematically well-defined structure, someday even rigorously so, and eventually we will understand it.

I believe that it will be largely understood and alternatives to string theory either found or ruled out on the 20-50 year timescale. Thus, although we must always be mindful of the current limits of knowledge, it is reasonable to think that the structures that have been discovered so far will be a significant portion of the final picture.

Summary

To summarize the status of the string landscape as a "top-down" concept which leads to candidate vacua, EFT, and predictions,

- Classification and calculational techniques are moderately well developed but by no means a mature subject.
- In addition people have suggested other relevant classes of 4d vacua (*e.g.* nonsupersymmetric), and in studying cosmology one may need to understand vacua with other numbers of large dimensions. So this is definitely a work in progress.
- On the other hand, the string landscape is a mathematically well-defined structure, someday even rigorously so, and eventually we will understand it.

I believe that it will be largely understood and alternatives to string theory either found or ruled out on the 20-50 year timescale. Thus, although we must always be mindful of the current limits of knowledge, it is reasonable to think that the structures that have been discovered so far will be a significant portion of the final picture.

Summary

To summarize the status of the string landscape as a "top-down" concept which leads to candidate vacua, EFT, and predictions,

- Classification and calculational techniques are moderately well developed but by no means a mature subject.
- In addition people have suggested other relevant classes of 4d vacua (*e.g.* nonsupersymmetric), and in studying cosmology one may need to understand vacua with other numbers of large dimensions. So this is definitely a work in progress.
- On the other hand, the string landscape is a mathematically well-defined structure, someday even rigorously so, and eventually we will understand it.

I believe that it will be largely understood and alternatives to string theory either found or ruled out on the 20-50 year timescale. Thus, although we must always be mindful of the current limits of knowledge, it is reasonable to think that the structures that have been discovered so far will be a significant portion of the final picture.

Interactions between theoretical physics, computer science and statistics

Scientific computation was one of the first important applications of computers and still drives many developments, for example quantum computation and supercomputing.

There have also been important conceptual interactions between the fields: to list a few,

- Physics of computation (Landauer, Fredkin, Bennett, Feynman, ...)
- Neural networks (Hopfield, Hinton, Sejnowski, ...)
- Simulated annealing, Monte Carlo algorithms for statistics
- Statistical physics - combinatorial optimization (Parisi, Mézard, ...)
- Matrix models and statistics (Marčenko-Pastur, Tracy-Widom, ...)
- Quantum information, entanglement and space-time (ongoing)

One direction of much current interest is to develop theories of deep learning. Relevant theoretical physics includes:

- Random matrix theory. For example, speed of optimization is governed by the eigenvalues of the Hessian of the objective function. These can be computed for complex feedforward networks in the large N limit using free probability theory – *e.g.* Ganguli, Pennington *et al*. 2017.
- Tensor networks. Networks which are not fully connected have internal structure which can sometimes be usefully reformulated by decomposing into tensors, *e.g.* Cohen, Sharir and Shashua 2015. Many people have speculated that tensor network formulations of the renormalization group (*e.g.* MERA) might be used to understand deep learning.

Interdisciplinary computational mathematics

Understanding string theory and making testable predictions is a long-term project: we might find success tomorrow, or in a decade, or in the distant future.

However long this takes, we can make important contributions to human knowledge along the way. For example, this is the spirit behind the math-physics interface: string theory "knows more mathematics than we do" and its study has led to the development of a large quantity of new mathematics: mirror symmetry (enumerative formulas, homological mirror symmetry, Bridgeland stability), new invariants (Gromov-Witten, Donaldson-Thomas, Gopakumar-Vafa, *etc.*), topological field theory and topological quantum gravity, *etc.*.

The contributions of string theorists are generally not rigorous mathematics but rather conjectures, new relations, new connections, and new questions which are developed in a loose collaboration between physicists and mathematicians.

Let us make the case that the exploration of the string landscape should also be done in collaboration with computer scientists and computational mathematicians. To summarize the arguments:

- String compactification uses a wide variety of mathematical results which are of central interest to mathematicians and mathematical scientists.
- Any systematic approach to its study, especially one which involves data science, will require making databases of string vacua.
- This requires having computational implementations of many mathematical structures.
- The math itself is intricate and computational implementations will be equally so. Especially, the work to validate and maintain them as computational and other technologies develop, and as our conceptual framework improves, exceeds what most academic collaborations can sustain.

- Fortunately these mathematical concepts and their computational implementations are of interest to larger communities. Finite group theory is of interest to all mathematical scientists, as are many statistical constructs. Lattices are of central interest in coding theory. Algebraic geometry and automorphic functions are more the province of pure mathematicians, but again there is a sizable community. There are applications of algebraic geometry to statistics (see work of Sturmfels *et al*) and theoretical computer science (see Mulmuley and Sohoni, many other works). Yau and others have advocated uses of Kähler geometry and moduli spaces in applied mathematics.
- There are also major computational challenges, not just in speed of algorithms and storage of large datasets, but in organizing a very large database of formal knowledge in a distributed and collaborative way. If there were a Wikipedia of computational mathematics, we could build on it. Since there is not, there should be a project to develop one. This is beyond what anyone knows how to do, but people are starting to think along these lines.

String theorists have made major contributions to computational mathematics: to list a few, the Kreuzer-Skarke database of CY_3's, the recent improvements of Altman *et al*, the recent list of $CICY_4$'s developed by Gray *et al*, and many, many more specialized works.

To some extent these contributions have been integrated into larger computer algebra systems (CAS) such as Sage, the largest and most actively developed open source CAS. A CAS provides both a platform to simplify writing mathematical software and a way to combine the many ingredients needed to study string compactifications.

Much relevant computational mathematics has been developed by other communities, for example computational group theory and representation theory, which for finite groups is in GAP (and in Sage).

Another area which will become valuable in the near term is automatic theorem verification and proof. The best developed systems for this are Coq and more recently Lean. Being able to verify our mathematical definitions would be of great help. The mathematics community is getting interested in this, see for example the "Big Proof" workshop at the INI and the "Formal Abstracts" initiative.

String theorists have made major contributions to computational mathematics: to list a few, the Kreuzer-Skarke database of CY_3's, the recent improvements of Altman *et al*, the recent list of $CICY_4$'s developed by Gray *et al*, and many, many more specialized works.

To some extent these contributions have been integrated into larger computer algebra systems (CAS) such as Sage, the largest and most actively developed open source CAS. A CAS provides both a platform to simplify writing mathematical software and a way to combine the many ingredients needed to study string compactifications.

Much relevant computational mathematics has been developed by other communities, for example computational group theory and representation theory, which for finite groups is in GAP (and in Sage). Another area which will become valuable in the near term is automatic theorem verification and proof. The best developed systems for this are Coq and more recently Lean. Being able to verify our mathematical definitions would be of great help. The mathematics community is getting interested in this, see for example the "Big Proof" workshop at the INI and the "Formal Abstracts" initiative.

String theorists have made major contributions to computational mathematics: to list a few, the Kreuzer-Skarke database of CY_3's, the recent improvements of Altman *et al*, the recent list of $CICY_4$'s developed by Gray *et al*, and many, many more specialized works.

To some extent these contributions have been integrated into larger computer algebra systems (CAS) such as Sage, the largest and most actively developed open source CAS. A CAS provides both a platform to simplify writing mathematical software and a way to combine the many ingredients needed to study string compactifications.

Much relevant computational mathematics has been developed by other communities, for example computational group theory and representation theory, which for finite groups is in GAP (and in Sage).

Another area which will become valuable in the near term is automatic theorem verification and proof. The best developed systems for this are Coq and more recently Lean. Being able to verify our mathematical definitions would be of great help. The mathematics community is getting interested in this, see for example the "Big Proof" workshop at the INI and the "Formal Abstracts" initiative.

String theorists have made major contributions to computational mathematics: to list a few, the Kreuzer-Skarke database of CY_3's, the recent improvements of Altman *et al*, the recent list of $CICY_4$'s developed by Gray *et al*, and many, many more specialized works.

To some extent these contributions have been integrated into larger computer algebra systems (CAS) such as Sage, the largest and most actively developed open source CAS. A CAS provides both a platform to simplify writing mathematical software and a way to combine the many ingredients needed to study string compactifications.

Much relevant computational mathematics has been developed by other communities, for example computational group theory and representation theory, which for finite groups is in GAP (and in Sage).

Another area which will become valuable in the near term is automatic theorem verification and proof. The best developed systems for this are Coq and more recently Lean. Being able to verify our mathematical definitions would be of great help. The mathematics community is getting interested in this, see for example the "Big Proof" workshop at the INI and the "Formal Abstracts" initiative.

While many of the mathematical definitions and algorithms we need to construct string vacua have been implemented, in my opinion there is a crucial missing ingredient – namely, a way to integrate them into larger projects, which is

- modular – continuous integration allowing changes in any component or even in the underlying definitions while minimizing the changes required to other components
- efficient – we have moderately large datasets with millions of items
- distributed – projects will generally be collaborative
- archival – facilitates publication and archiving of papers and datasets
- open source – not dependent on proprietary software

Many areas of scientific computation and open source software development have these requirements, so this is another problem which our community should help to define, but should not have to solve on its own.

Conclusions

At the moment, the most important problems for string theorists to solve are still physical:

- Establish existence of a metastable de Sitter vacuum to everyone's satisfaction.
- Find ways to observe the hidden sectors: amount of dark matter, axion superradiance, etc.
- Understand the dynamics that leads to four large dimensions.
- Understand what tunneling to AdS means in eternal inflation.
- Find a believable toy model of the entire landscape and study issues such as simple/complex and nature of measure factors.
- Make conjectures about preferred initial conditions, perhaps non-geometrical.

But in the longer term, string/M theorists will need computer/data science to make progress. Hopefully the benefits will be mutual. Clearly the main thing we need are interesting datasets of vacua/EFTs and tools to work with them. Then, we can hope to find interesting patterns in the distribution of topological invariants, in couplings or in observables. Our longer term goal is to do computations on the "entire landscape" which are beyond analytic techniques, to decide what are the likely predictions of string theory.

Here are some more near term ways to bring the fields together:

- Agree on a standard format for EFTs and other commonly used physical constructs.
- Likewise for the most important math (Lie groups and representations, CY manifold data)
- Make contact with all of the active developers of mathematical software and discuss our needs.
- Standard nomenclature for individual string compactifications?
- Develop statistics which compare the completeness or representativeness of ensembles of string vacua.

But in the longer term, string/M theorists will need computer/data science to make progress. Hopefully the benefits will be mutual. Clearly the main thing we need are interesting datasets of vacua/EFTs and tools to work with them. Then, we can hope to find interesting patterns in the distribution of topological invariants, in couplings or in observables. Our longer term goal is to do computations on the "entire landscape" which are beyond analytic techniques, to decide what are the likely predictions of string theory.

Here are some more near term ways to bring the fields together:

- Agree on a standard format for EFTs and other commonly used physical constructs.
- Likewise for the most important math (Lie groups and representations, CY manifold data)
- Make contact with all of the active developers of mathematical software and discuss our needs.
- Standard nomenclature for individual string compactifications?
- Develop statistics which compare the completeness or representativeness of ensembles of string vacua