Thoughts on $(2,0)$ theory

Michael R. Douglas

Simons Center / Stony Brook University

Strings 2013, Sogang University, Seoul, Korea

Abstract

We discuss the $(2,0)$ superconformal field theory in six dimensions and its relation to $D = 5$ maximally supersymmetric Yang-Mills theory, mostly based on arXiv:1012.2880.
We also report on arXiv:1210.7709 with Zvi Bern, John Joseph Carrasco, Lance Dixon, Matt von Hippel and Henrik Johansson in which we showed that $D = 5$ MSYM diverges at six loops.

Outline

Inspired by string/M theory and AdS/CFT, many superconformal field theories have been discovered and studied. The most mysterious and arguably the most important is the $(2,0)$ theory in six dimensions (Witten, 1995). There is a free or "abelian" $(2,0)$ theory, a field theory of a self-dual tensor (3 physical modes), 5 scalars and 8 fermions.

There are also "nonabelian" $(2,0)$ theories, with various indirect definitions and relations to other superconformal theories:

- Type IIb string theory on a four-dimensional ADE singularity, leading to an ADE classification of these theories.
- N coincident M5-branes, leading to the A_{N-1} theory.
- The large N limit is dual to M theory on $AdS_7 \times S^4$.
- After compactification on T^2, one gets $N = 4$ maximally supersymmetric Yang-Mills theory, with complex gauge coupling τ equal to the complex modulus of T^2. This gives S-duality a geometric explanation.
- Twisted compactifications on manifolds with $d = 2, 3, 4$ lead to new superconformal theories in $D = 4, 3, 2$.

Inspired by string/M theory and AdS/CFT, many superconformal field theories have been discovered and studied. The most mysterious and arguably the most important is the $(2,0)$ theory in six dimensions (Witten, 1995). There is a free or "abelian" $(2,0)$ theory, a field theory of a self-dual tensor (3 physical modes), 5 scalars and 8 fermions. There are also "nonabelian" $(2,0)$ theories, with various indirect definitions and relations to other superconformal theories:

- Type IIb string theory on a four-dimensional ADE singularity, leading to an ADE classification of these theories.
- N coincident M5-branes, leading to the A_{N-1} theory.
- The large N limit is dual to M theory on $AdS_7 \times S^4$.
- After compactification on T^2, one gets $N = 4$ maximally supersymmetric Yang-Mills theory, with complex gauge coupling τ equal to the complex modulus of T^2. This gives S-duality a geometric explanation.
- Twisted compactifications on manifolds with $d = 2, 3, 4$ lead to new superconformal theories in $D = 4, 3, 2$.

However there is no satisfactory Lagrangian, and since the theory has no dimensionless coupling, no argument that there must be one, nor is any other usable microscopic definition known.

Here are some of the ideas which have been tried:

- As discussed here by John Schwarz, on the Coulomb branch there are BPS strings, with tension proportional to differences of scalar vevs $\phi_i - \phi_j$. In the unbroken limit $\phi_i \to \phi_j$, these might lead to "tensionless strings" as fundamental degrees of freedom.
- One can conjecturally define M theory in the light-cone frame as the large N limit of D0-branes in IIa theory – BFSS "Matrix theory." This idea also leads to a definition of the A_N and D_N $(2,0)$ theories as large N limits of D0-D4 systems.
- One might start with the theory compactified on S^1, T^2 or possibly other manifolds, add in "Kaluza-Klein" and other extra degrees of freedom of the $D = 6$ theory, and take the large volume limit.

However there is no satisfactory Lagrangian, and since the theory has no dimensionless coupling, no argument that there must be one, nor is any other usable microscopic definition known.
Here are some of the ideas which have been tried:

- As discussed here by John Schwarz, on the Coulomb branch there are BPS strings, with tension proportional to differences of scalar vevs $\phi_i - \phi_j$. In the unbroken limit $\phi_i \to \phi_j$, these might lead to "tensionless strings" as fundamental degrees of freedom.
- One can conjecturally define M theory in the light-cone frame as the large N limit of D0-branes in IIa theory – BFSS "Matrix theory." This idea also leads to a definition of the A_N and D_N $(2,0)$ theories as large N limits of D0-D4 systems.
- One might start with the theory compactified on S^1, T^2 or possibly other manifolds, add in "Kaluza-Klein" and other extra degrees of freedom of the $D = 6$ theory, and take the large volume limit.

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

What do we think we know about $(2,0)$ theory, beyond its defining properties and BPS spectrum?

- It is a local field theory, with local correlation functions, which can be defined in any (nonsingular) space-time with a fixed metric.
- It has no dimensionless parameters, and no relevant operators (but scales can be introduced by going on the Coulomb branch).
- The A_N and D_N theories have $O(N^3)$ degrees of freedom, as measured by the free energy, or by the Weyl anomaly (Henningson and Skenderis 1998). Maxwell and Sethi 1204.2002 have a (complicated) field theory argument that the conformal anomaly is $O(N^3)$.
- The theory involves extended objects – not only do they exist, but if we remove them the theory becomes non-interacting.

Extended objects in a conformal theory must be very strange indeed, as they can have no preferred size – they are "scale-free." Why do we believe the last claim?

There is a very general argument that continuum theories of weakly interacting particles cannot exist in more than four space-time dimensions. The paths which contribute to the path integral of a free quantum particle have fractal dimension 2 (this is perhaps more familiar in the statistical mechanical analog of Brownian motion). Two generic submanifolds of dimension d, embedded in a D-dimensional space-time, would be expected to intersect in a manifold of dimension $2d - D$. Thus, for $D > 4$, particle paths will not intersect (more precisely the probability or amplitude of their intersection is zero).

In a cutoff theory, the cutoff will spread out the interaction and evade this argument, but then the interactions will go to zero in the IR as a power of the cutoff. This has even been made rigorous in some (rather restrictive) cases (Aizenman, 1981).

Could this be evaded by strongly interacting particles, perhaps because any picture in terms of paths involves infinitely many particles? Maybe – but if these particles each carry an infinitesimal portion of the momentum of a process, then it looks like they are forming an extended object. Would be nice to make this sharp.

There is a very general argument that continuum theories of weakly interacting particles cannot exist in more than four space-time dimensions. The paths which contribute to the path integral of a free quantum particle have fractal dimension 2 (this is perhaps more familiar in the statistical mechanical analog of Brownian motion). Two generic submanifolds of dimension d, embedded in a D-dimensional space-time, would be expected to intersect in a manifold of dimension $2d - D$. Thus, for $D > 4$, particle paths will not intersect (more precisely the probability or amplitude of their intersection is zero).
In a cutoff theory, the cutoff will spread out the interaction and evade this argument, but then the interactions will go to zero in the IR as a power of the cutoff. This has even been made rigorous in some (rather restrictive) cases (Aizenman, 1981).
Could this be evaded by strongly interacting particles, perhaps because any picture in terms of paths involves infinitely many particles? Maybe – but if these particles each carry an infinitesimal portion of the momentum of a process, then it looks like they are forming an extended object. Would be nice to make this sharp.

There is a very general argument that continuum theories of weakly interacting particles cannot exist in more than four space-time dimensions. The paths which contribute to the path integral of a free quantum particle have fractal dimension 2 (this is perhaps more familiar in the statistical mechanical analog of Brownian motion). Two generic submanifolds of dimension d, embedded in a D-dimensional space-time, would be expected to intersect in a manifold of dimension $2d - D$. Thus, for $D > 4$, particle paths will not intersect (more precisely the probability or amplitude of their intersection is zero).
In a cutoff theory, the cutoff will spread out the interaction and evade this argument, but then the interactions will go to zero in the IR as a power of the cutoff. This has even been made rigorous in some (rather restrictive) cases (Aizenman, 1981).
Could this be evaded by strongly interacting particles, perhaps because any picture in terms of paths involves infinitely many particles? Maybe – but if these particles each carry an infinitesimal portion of the momentum of a process, then it looks like they are forming an extended object. Would be nice to make this sharp.

If we grant that $(2,0)$ theory involves extended objects, what are they? The original answer to this question, motivated by the string theory arguments, is that they are tensionless strings – literally tensionless in the symmetry limit, and effectively tensionless at energies high compared to $|\phi_i - \phi_j|$.

There are actions for tensionless strings, starting with Schild 1977, and a modest literature which began before the proposal of $(2,0)$ theory, in good part motivated by the desire to understand the high energy limit of string scattering amplitudes, which has many interesting properties (Gross and Mende 1988; Amati, Ciafaloni, and Veneziano 1988; Sundborg 1988; ...). The simplest action (Lindström Sundborg Theodoridis 1991) is

$$S = \int d^2\sigma \; V^\alpha V^\beta D_\alpha X \cdot D_\beta X + \lambda X^2 \tag{1}$$

where X embeds the string into the lightcone of $\mathbb{R}^{D,2}$, there is a gauged scale invariance, and V is a 2d vector (the world-sheet metric is rank 1).

If we grant that $(2,0)$ theory involves extended objects, what are they? The original answer to this question, motivated by the string theory arguments, is that they are tensionless strings – literally tensionless in the symmetry limit, and effectively tensionless at energies high compared to $|\phi_i - \phi_j|$.

There are actions for tensionless strings, starting with Schild 1977, and a modest literature which began before the proposal of $(2,0)$ theory, in good part motivated by the desire to understand the high energy limit of string scattering amplitudes, which has many interesting properties (Gross and Mende 1988; Amati, Ciafaloni, and Veneziano 1988; Sundborg 1988; ...). The simplest action (Lindström Sundborg Theodoridis 1991) is

$$S = \int d^2\sigma \; V^\alpha V^\beta D_\alpha X \cdot D_\beta X + \lambda X^2 \tag{1}$$

where X embeds the string into the lightcone of $\mathbb{R}^{D,2}$, there is a gauged scale invariance, and V is a 2d vector (the world-sheet metric is rank 1).

So far most work is inconclusive, while the few conclusive works are negative – for example Gustafsson *et al* hep-th/9410143 showed that the action above can only be quantized in $D = 2$. Any string description would also have to address the problem that strings normally do not allow defining local correlation functions.

Another idea is that $(2, 0)$ theory is simpler in twistor space or some other formulation which is not manifestly local, evading the arguments above – *e.g.* see Mason and Reid-Edwards 1212.6173 and Sämann and Wolf 1205.3108.

Finally, there is the idea that these concrete classical descriptions are not worth developing, because the theory is strongly coupled anyways. Despite this obstacle, one might find particular amplitudes which can be understood perturbatively, one might find ways to resum perturbation theory, or one might work with a Hamiltonian formulation and explicit wave functions. There is no reason to be defeatist!

So far most work is inconclusive, while the few conclusive works are negative – for example Gustafsson *et al* hep-th/9410143 showed that the action above can only be quantized in $D = 2$. Any string description would also have to address the problem that strings normally do not allow defining local correlation functions.

Another idea is that $(2, 0)$ theory is simpler in twistor space or some other formulation which is not manifestly local, evading the arguments above – *e.g.* see Mason and Reid-Edwards 1212.6173 and Sämann and Wolf 1205.3108.

Finally, there is the idea that these concrete classical descriptions are not worth developing, because the theory is strongly coupled anyways. Despite this obstacle, one might find particular amplitudes which can be understood perturbatively, one might find ways to resum perturbation theory, or one might work with a Hamiltonian formulation and explicit wave functions. There is no reason to be defeatist!

So far most work is inconclusive, while the few conclusive works are negative – for example Gustafsson *et al* hep-th/9410143 showed that the action above can only be quantized in $D = 2$. Any string description would also have to address the problem that strings normally do not allow defining local correlation functions.

Another idea is that $(2, 0)$ theory is simpler in twistor space or some other formulation which is not manifestly local, evading the arguments above – *e.g.* see Mason and Reid-Edwards 1212.6173 and Sämann and Wolf 1205.3108.

Finally, there is the idea that these concrete classical descriptions are not worth developing, because the theory is strongly coupled anyways. Despite this obstacle, one might find particular amplitudes which can be understood perturbatively, one might find ways to resum perturbation theory, or one might work with a Hamiltonian formulation and explicit wave functions. There is no reason to be defeatist!

The simplest relation between $(2, 0)$ theory and a conventional field theory is to compactify on S^1. This leads to $D = 5$ maximally supersymmetric Yang-Mills – the self-dual tensor reduces to an ordinary vector boson in $D = 5$ (the extra components H_{IJK} are dual to $H_{5LM} = F_{LM}$). Non simply laced gauge groups in $D = 5$ come from twisted boundary conditions on S^1.

The Yang-Mills coupling g_5^2 in $D = 5$ has dimensions of length. Thus, as suggested by dimensional analysis, and confirmed by BPS arguments, one has

$$g_5^2 = R_5$$

in terms of the radius of S^1. Thus the prefactor $1/g_5^2$ of the action is not what would come from standard KK reduction and $\int dx^5$.

This theory is of course nonrenormalizable. The perturbative expansion is controlled by the effective dimensionless coupling $g_5^2 E$, where E is the energy scale of a process. At low energies, the expansion should be good, while for $E \sim 1/g_5^2$ it will break down.

The simplest relation between $(2, 0)$ theory and a conventional field theory is to compactify on S^1. This leads to $D = 5$ maximally supersymmetric Yang-Mills – the self-dual tensor reduces to an ordinary vector boson in $D = 5$ (the extra components H_{IJK} are dual to $H_{5LM} = F_{LM}$). Non simply laced gauge groups in $D = 5$ come from twisted boundary conditions on S^1.
The Yang-Mills coupling g_5^2 in $D = 5$ has dimensions of length. Thus, as suggested by dimensional analysis, and confirmed by BPS arguments, one has

$$g_5^2 = R_5$$

in terms of the radius of S^1. Thus the prefactor $1/g_5^2$ of the action is not what would come from standard KK reduction and $\int dx^5$.

This theory is of course nonrenormalizable. The perturbative expansion is controlled by the effective dimensionless coupling $g_5^2 E$, where E is the energy scale of a process. At low energies, the expansion should be good, while for $E \sim 1/g_5^2$ it will break down.

The simplest relation between $(2, 0)$ theory and a conventional field theory is to compactify on S^1. This leads to $D = 5$ maximally supersymmetric Yang-Mills – the self-dual tensor reduces to an ordinary vector boson in $D = 5$ (the extra components H_{IJK} are dual to $H_{5LM} = F_{LM}$). Non simply laced gauge groups in $D = 5$ come from twisted boundary conditions on S^1.
The Yang-Mills coupling g_5^2 in $D = 5$ has dimensions of length. Thus, as suggested by dimensional analysis, and confirmed by BPS arguments, one has

$$g_5^2 = R_5$$

in terms of the radius of S^1. Thus the prefactor $1/g_5^2$ of the action is not what would come from standard KK reduction and $\int dx^5$.

This theory is of course nonrenormalizable. The perturbative expansion is controlled by the effective dimensionless coupling $g_5^2 E$, where E is the energy scale of a process. At low energies, the expansion should be good, while for $E \sim 1/g_5^2$ it will break down.

The BPS objects are derived from the standard solitons, instantons and dyons of Yang-Mills theory, which have new interpretations in $D = 5$:

- The self-dual solutions lead to particles. The sector with topological charge n has lowest energy $n/g_5^2 = n/R_5$, corresponding to a state with KK momentum n.
- The 't Hooft-Polyakov solution (in the Coulomb phase) becomes an infinite string. This is the $D = 5$ counterpart of the $D = 6$ string or M theory M2-brane stretched between M5-branes.
- These monopole strings can form string junctions, which are $1/4$ BPS states. They have $O(N^3)$ multiplicity and might be relevant for explaining the N^3 free energy (Lee and Yee 2007, Bolognesi and Lee 2011).

The self-dual solutions can be 1/2 BPS and were identified as KK excitations of the elementary $(2, 0)$ multiplets in the early works (Berkooz, Rozali, Seiberg 1997) and also have this interpretation in the Matrix theory realization (Berkooz and Douglas 1996). However they have never been well understood. Self-dual solutions have many additional moduli, including a scale size ρ, which do not fit with a conventional particle interpretation.

Some of these puzzles can be solved more easily on the Coulomb branch. For example, Lambert *et al* 2010 found 1/4 BPS dyonic instantons, which correspond to self-dual strings wrapped on S^1, and have no scale size modulus. However as we restore symmetry they blow up.

It seems that the interpretation of these objects and specifically the scale size is a key point in understanding these theories. We will suggest that they are in fact the scale-free extended objects we argued must exist in a nontrivial $D > 4$ CFT.

The self-dual solutions can be 1/2 BPS and were identified as KK excitations of the elementary $(2, 0)$ multiplets in the early works (Berkooz, Rozali, Seiberg 1997) and also have this interpretation in the Matrix theory realization (Berkooz and Douglas 1996). However they have never been well understood. Self-dual solutions have many additional moduli, including a scale size ρ, which do not fit with a conventional particle interpretation.

Some of these puzzles can be solved more easily on the Coulomb branch. For example, Lambert *et al* 2010 found 1/4 BPS dyonic instantons, which correspond to self-dual strings wrapped on S^1, and have no scale size modulus. However as we restore symmetry they blow up.

It seems that the interpretation of these objects and specifically the scale size is a key point in understanding these theories. We will suggest that they are in fact the scale-free extended objects we argued must exist in a nontrivial $D > 4$ CFT.

The self-dual solutions can be 1/2 BPS and were identified as KK excitations of the elementary $(2, 0)$ multiplets in the early works (Berkooz, Rozali, Seiberg 1997) and also have this interpretation in the Matrix theory realization (Berkooz and Douglas 1996). However they have never been well understood. Self-dual solutions have many additional moduli, including a scale size ρ, which do not fit with a conventional particle interpretation.

Some of these puzzles can be solved more easily on the Coulomb branch. For example, Lambert *et al* 2010 found 1/4 BPS dyonic instantons, which correspond to self-dual strings wrapped on S^1, and have no scale size modulus. However as we restore symmetry they blow up.

It seems that the interpretation of these objects and specifically the scale size is a key point in understanding these theories. We will suggest that they are in fact the scale-free extended objects we argued must exist in a nontrivial $D > 4$ CFT.

One can make various hypotheses about the precise relation between the two theories. The first question to address is the expected UV divergences in $D = 5$ MSYM. By power counting, one expects an ℓ-loop amplitude in pure $D = 5$ YM to go as

$$\mathcal{A}_\ell \sim g_5^{2\ell-2}\Lambda^\ell$$

where Λ is a UV cutoff. These divergences are mitigated in MSYM but, at least according to superspace arguments, are still expected (more on this point below).

On the other hand, $(2, 0)$ theory has no scale, and no UV divergences. Thus it provides a UV completion of $D = 5$ super Yang-Mills and an effective cutoff Λ.

While there are many examples of UV completions of nonrenormalizable theories, this one seems somehow different:

- The Fermi theory of weak interactions has as UV completion the Weinberg-Salam model (and ultimately a GUT). In this example, the underlying theory has two dimensionless gauge couplings.
- More generally, given a nonrenormalizable theory with a coupling $g \sim$ (length), and a cutoff Λ, one expects to get a dimensionless coupling $g\Lambda$ in the UV completion. Another example is the string theory UV completion of $D = 10$ supergravity.
- A different example is the M theory UV completion of $D = 11$ supergravity. In this case, we believe the cutoff $\Lambda \sim 1/l_{p11}$, and is fixed in terms of the coupling.

This last example seems most relevant, and thus we postulate

$$\Lambda = 1/g_5^2 = 1/R_5$$

in $D = 5$ MSYM with the $(2, 0)$ completion. Like M theory, there is no dimensionless coupling. A significant difference is that in $(2, 0)$ theory, the cutoff only arises after we compactify.

While there are many examples of UV completions of nonrenormalizable theories, this one seems somehow different:

- The Fermi theory of weak interactions has as UV completion the Weinberg-Salam model (and ultimately a GUT). In this example, the underlying theory has two dimensionless gauge couplings.
- More generally, given a nonrenormalizable theory with a coupling $g \sim$ (length), and a cutoff Λ, one expects to get a dimensionless coupling $g\Lambda$ in the UV completion. Another example is the string theory UV completion of $D = 10$ supergravity.
- A different example is the M theory UV completion of $D = 11$ supergravity. In this case, we believe the cutoff $\Lambda \sim 1/l_{p11}$, and is fixed in terms of the coupling.

This last example seems most relevant, and thus we postulate

$$\Lambda = 1/g_5^2 = 1/R_5$$

in $D = 5$ MSYM with the $(2, 0)$ completion. Like M theory, there is no dimensionless coupling. A significant difference is that in $(2, 0)$ theory, the cutoff only arises after we compactify.

While there are many examples of UV completions of nonrenormalizable theories, this one seems somehow different:

- The Fermi theory of weak interactions has as UV completion the Weinberg-Salam model (and ultimately a GUT). In this example, the underlying theory has two dimensionless gauge couplings.
- More generally, given a nonrenormalizable theory with a coupling $g \sim$ (length), and a cutoff Λ, one expects to get a dimensionless coupling $g\Lambda$ in the UV completion. Another example is the string theory UV completion of $D = 10$ supergravity.
- A different example is the M theory UV completion of $D = 11$ supergravity. In this case, we believe the cutoff $\Lambda \sim 1/l_{p11}$, and is fixed in terms of the coupling.

This last example seems most relevant, and thus we postulate

$$\Lambda = 1/g_5^2 = 1/R_5$$

in $D = 5$ MSYM with the $(2, 0)$ completion. Like M theory, there is no dimensionless coupling. A significant difference is that in $(2, 0)$ theory, the cutoff only arises after we compactify.

While there are many examples of UV completions of nonrenormalizable theories, this one seems somehow different:

- The Fermi theory of weak interactions has as UV completion the Weinberg-Salam model (and ultimately a GUT). In this example, the underlying theory has two dimensionless gauge couplings.
- More generally, given a nonrenormalizable theory with a coupling $g \sim$ (length), and a cutoff Λ, one expects to get a dimensionless coupling $g\Lambda$ in the UV completion. Another example is the string theory UV completion of $D = 10$ supergravity.
- A different example is the M theory UV completion of $D = 11$ supergravity. In this case, we believe the cutoff $\Lambda \sim 1/l_{p11}$, and is fixed in terms of the coupling.

This last example seems most relevant, and thus we postulate

$$\Lambda = 1/g_5^2 = 1/R_5$$

in $D = 5$ MSYM with the $(2, 0)$ completion. Like M theory, there is no dimensionless coupling. A significant difference is that in $(2, 0)$ theory, the cutoff only arises after we compactify.

To be more precise, we consider exact correlation functions or S-matrix elements in compactified $(2, 0)$ theory. These will be functions of an overall energy scale E and will cross over from $D = 5$ behavior at $E << 1/R_5$ to $(2, 0)$ behavior at $E >> 1/R_5$. The crossover scale defines a cutoff $\Lambda \sim 1/R_5 \sim 1/g_5^2$.

Where does this cutoff come from in $D = 5$ terms?

One can make various hypotheses:

1. Contrary to expectations, there are no UV divergences in standard $D = 5$ MSYM – it is perturbatively finite. This turns out not to be the case.
2. New degrees of freedom come in at the energy $E \sim 1/g_5^2$ and cancel the UV divergences, providing an effective cutoff. They might also explain the $O(N^3)$ free energy.
3. There are no new degrees of freedom, but somehow resummation of the perturbative series eliminates the divergences.
4. The whole discussion is based on a misconception, because S-matrix elements and other observables do not have a consistent definition between the two theories.

To be more precise, we consider exact correlation functions or S-matrix elements in compactified $(2, 0)$ theory. These will be functions of an overall energy scale E and will cross over from $D = 5$ behavior at $E << 1/R_5$ to $(2, 0)$ behavior at $E >> 1/R_5$. The crossover scale defines a cutoff $\Lambda \sim 1/R_5 \sim 1/g_5^2$.

Where does this cutoff come from in $D = 5$ terms?

One can make various hypotheses:

1. Contrary to expectations, there are no UV divergences in standard $D = 5$ MSYM – it is perturbatively finite. This turns out not to be the case.
2. New degrees of freedom come in at the energy $E \sim 1/g_5^2$ and cancel the UV divergences, providing an effective cutoff. They might also explain the $O(N^3)$ free energy.
3. There are no new degrees of freedom, but somehow resummation of the perturbative series eliminates the divergences.
4. The whole discussion is based on a misconception, because S-matrix elements and other observables do not have a consistent definition between the two theories.

To be more precise, we consider exact correlation functions or S-matrix elements in compactified $(2, 0)$ theory. These will be functions of an overall energy scale E and will cross over from $D = 5$ behavior at $E << 1/R_5$ to $(2, 0)$ behavior at $E >> 1/R_5$. The crossover scale defines a cutoff $\Lambda \sim 1/R_5 \sim 1/g_5^2$.

Where does this cutoff come from in $D = 5$ terms?

One can make various hypotheses:

1. Contrary to expectations, there are no UV divergences in standard $D = 5$ MSYM – it is perturbatively finite. This turns out not to be the case.
2. New degrees of freedom come in at the energy $E \sim 1/g_5^2$ and cancel the UV divergences, providing an effective cutoff. They might also explain the $O(N^3)$ free energy.
3. There are no new degrees of freedom, but somehow resummation of the perturbative series eliminates the divergences.
4. The whole discussion is based on a misconception, because S-matrix elements and other observables do not have a consistent definition between the two theories.

To be more precise, we consider exact correlation functions or S-matrix elements in compactified $(2, 0)$ theory. These will be functions of an overall energy scale E and will cross over from $D = 5$ behavior at $E << 1/R_5$ to $(2, 0)$ behavior at $E >> 1/R_5$. The crossover scale defines a cutoff $\Lambda \sim 1/R_5 \sim 1/g_5^2$.

Where does this cutoff come from in $D = 5$ terms?

One can make various hypotheses:

1. Contrary to expectations, there are no UV divergences in standard $D = 5$ MSYM – it is perturbatively finite. This turns out not to be the case.
2. New degrees of freedom come in at the energy $E \sim 1/g_5^2$ and cancel the UV divergences, providing an effective cutoff. They might also explain the $O(N^3)$ free energy.
3. There are no new degrees of freedom, but somehow resummation of the perturbative series eliminates the divergences.
4. The whole discussion is based on a misconception, because S-matrix elements and other observables do not have a consistent definition between the two theories.

To be more precise, we consider exact correlation functions or S-matrix elements in compactified $(2, 0)$ theory. These will be functions of an overall energy scale E and will cross over from $D = 5$ behavior at $E << 1/R_5$ to $(2, 0)$ behavior at $E >> 1/R_5$. The crossover scale defines a cutoff $\Lambda \sim 1/R_5 \sim 1/g_5^2$.

Where does this cutoff come from in $D = 5$ terms?

One can make various hypotheses:

1. Contrary to expectations, there are no UV divergences in standard $D = 5$ MSYM – it is perturbatively finite. This turns out not to be the case.
2. New degrees of freedom come in at the energy $E \sim 1/g_5^2$ and cancel the UV divergences, providing an effective cutoff. They might also explain the $O(N^3)$ free energy.
3. There are no new degrees of freedom, but somehow resummation of the perturbative series eliminates the divergences.
4. The whole discussion is based on a misconception, because S-matrix elements and other observables do not have a consistent definition between the two theories.

Indeed, while one can usually define the S-matrix of a superconformal theory (as for $D = 4$ MSYM), this is not yet clear for the $(2, 0)$ theory. For example, the $D = 6$ 3-point S-matrix element for self-dual tensor multiplets vanishes for symmetry reasons (Huang and Lipstein, 1004.4735). Although one can get interactions between two tensors and a higher spin field, in fact these can only be supergravity amplitudes (Czech, Huang and Rozali, 1110.2791).

Although there may well be subtleties in defining the $D = 6$ and even the $D = 5$ S-matrix in the unbroken phase, these can be dealt with by working on the Coulomb branch, on which the theory is free at low energies. The question of UV divergences can be asked for low energy scattering as well, so we cannot answer it just in terms of subtleties in defining the S-matrix.

Indeed, while one can usually define the S-matrix of a superconformal theory (as for $D = 4$ MSYM), this is not yet clear for the $(2, 0)$ theory. For example, the $D = 6$ 3-point S-matrix element for self-dual tensor multiplets vanishes for symmetry reasons (Huang and Lipstein, 1004.4735). Although one can get interactions between two tensors and a higher spin field, in fact these can only be supergravity amplitudes (Czech, Huang and Rozali, 1110.2791).

Although there may well be subtleties in defining the $D = 6$ and even the $D = 5$ S-matrix in the unbroken phase, these can be dealt with by working on the Coulomb branch, on which the theory is free at low energies. The question of UV divergences can be asked for low energy scattering as well, so we cannot answer it just in terms of subtleties in defining the S-matrix.

Thus we are led to possibility 2, new degrees of freedom. There are three possibilities:

1. They are BPS and we can already see them in $D = 5$ MSYM.
2. They are finite size analogs of infinite size BPS objects, such as finite tensionless strings.
3. They are non-BPS and have no BPS analogs.

I have no ideas for the last possibility. In all three cases, we should explain why in the compactified theory, the new degrees of freedom only make their appearance at length scales $L \lesssim 1/\Lambda = R_5$.

Thus we are led to possibility 2, new degrees of freedom. There are three possibilities:

1. They are BPS and we can already see them in $D = 5$ MSYM.
2. They are finite size analogs of infinite size BPS objects, such as finite tensionless strings.
3. They are non-BPS and have no BPS analogs.

I have no ideas for the last possibility. In all three cases, we should explain why in the compactified theory, the new degrees of freedom only make their appearance at length scales $L \lesssim 1/\Lambda = R_5$.

Let us think about the second possibility. Consider the Coulomb branch with $|\phi_i - \phi_j| << 1/g_5^2$, then the monopole strings are very light, and it is plausible that these become infinite tensionless strings in the symmetry limit. What is their dynamics? Could they vibrate and intercommute, leading to finite loops of tensionless string?

Since $D = 5$ YM is very weakly coupled in this regime, we can study this question before taking the symmetry limit, on the Coulomb branch. The problem is very similar to that of vortex strings in $D = 4$, whose tension is $T = 2\pi\langle\phi\rangle^2$. Whlie one can make finite loops of vortex string, they are very unstable, and not important dynamically.

As for the monopole string, its size and energy densities are set by $|\phi_i - \phi_j|$, so it is not clear why its nature should change at the length scale g_5^2. Taking the symmetry limit only increases this length scale, so there is no reason to think it plays a role in the cutoff Λ. These seem to me good arguments against possibility 2.

Let us think about the second possibility. Consider the Coulomb branch with $|\phi_i - \phi_j| << 1/g_5^2$, then the monopole strings are very light, and it is plausible that these become infinite tensionless strings in the symmetry limit. What is their dynamics? Could they vibrate and intercommute, leading to finite loops of tensionless string?

Since $D = 5$ YM is very weakly coupled in this regime, we can study this question before taking the symmetry limit, on the Coulomb branch.

The problem is very similar to that of vortex strings in $D = 4$, whose tension is $T = 2\pi\langle\phi\rangle^2$. Whlie one can make finite loops of vortex string, they are very unstable, and not important dynamically.

As for the monopole string, its size and energy densities are set by $|\phi_i - \phi_j|$, so it is not clear why its nature should change at the length scale g_5^2. Taking the symmetry limit only increases this length scale, so there is no reason to think it plays a role in the cutoff Λ. These seem to me good arguments against possibility 2.

Let us think about the second possibility. Consider the Coulomb branch with $|\phi_i - \phi_j| << 1/g_5^2$, then the monopole strings are very light, and it is plausible that these become infinite tensionless strings in the symmetry limit. What is their dynamics? Could they vibrate and intercommute, leading to finite loops of tensionless string?
Since $D = 5$ YM is very weakly coupled in this regime, we can study this question before taking the symmetry limit, on the Coulomb branch.
The problem is very similar to that of vortex strings in $D = 4$, whose tension is $T = 2\pi\langle\phi\rangle^2$. Whlie one can make finite loops of vortex string, they are very unstable, and not important dynamically.
As for the monopole string, its size and energy densities are set by $|\phi_i - \phi_j|$, so it is not clear why its nature should change at the length scale g_5^2. Taking the symmetry limit only increases this length scale, so there is no reason to think it plays a role in the cutoff Λ. These seem to me good arguments against possibility 2.

To be more precise, there could be a description of $(2, 0)$ in terms of fundamental tensionless strings, and they might be good descriptions of $D = 6$ dynamics (at $L << R_5$), but at long distances their dynamics would have to be quite different than that of the monopole strings of $D = 5$ SYM. Thus they do not seem like a good way to think about the UV completion of the $D = 5$ theory.

Thus, we return to the first alternative. The instantonic particle has mass $1/R_5 = \Lambda$ so it naturally appears at the scale we need to cutoff the theory. A simple and attractive hypothesis is that it also contributes to loop amplitudes, and will change the UV divergences, perhaps cancelling them. To study this we need to develop the perturbation theory.

To be more precise, there could be a description of $(2, 0)$ in terms of fundamental tensionless strings, and they might be good descriptions of $D = 6$ dynamics (at $L << R_5$), but at long distances their dynamics would have to be quite different than that of the monopole strings of $D = 5$ SYM. Thus they do not seem like a good way to think about the UV completion of the $D = 5$ theory.

Thus, we return to the first alternative. The instantonic particle has mass $1/R_5 = \Lambda$ so it naturally appears at the scale we need to cutoff the theory. A simple and attractive hypothesis is that it also contributes to loop amplitudes, and will change the UV divergences, perhaps cancelling them. To study this we need to develop the perturbation theory.

Superspace arguments are more or less dimension independent and lead to the following:

- The YM action $\mathrm{tr}\ \mathrm{F}^2 + \ldots$ is not renormalized.
- The leading renormalization is of $\mathrm{tr}\ \mathrm{F}^4 + \ldots$, but this is only at one loop.
- Terms $\mathrm{tr}\ \mathrm{D}^2\mathrm{F}^4 + \ldots$ and higher order terms can be generated at all loop orders.

Since $\mathrm{D}^2\mathrm{F}^4$ is dimension 10, power counting predicts the first $\mathrm{D} = 5$ UV divergence at

$$(\mathrm{D} - 4)\ell = 10 - 4 = 6 \text{ loops}.$$

In Bern *et al* 1210.7709 we have verified that there is such a divergence, by explicit computation. If time permits we will discuss this later.

Superspace arguments are more or less dimension independent and lead to the following:

- The YM action $tr\ F^2 + \ldots$ is not renormalized.
- The leading renormalization is of $tr\ F^4 + \ldots$, but this is only at one loop.
- Terms $tr\ D^2 F^4 + \ldots$ and higher order terms can be generated at all loop orders.

Since $D^2 F^4$ is dimension 10, power counting predicts the first $D = 5$ UV divergence at

$$(D-4)\ell = 10 - 4 = 6 \text{ loops}.$$

In Bern *et al* 1210.7709 we have verified that there is such a divergence, by explicit computation. If time permits we will discuss this later.

It is hard to know what properties we should look for in $D = 5$ amplitudes, but we do know a fair amount about the $D = 4$ amplitudes obtained by compactification on T^2. Let the volume of the T^2 be L^2, then as $L \to 0$ we recover $N = 4$ SYM, while we expect corrections to come with positive powers of L. For example,

$$S_{eff} = \int \frac{1}{g_4^2} tr\ F^2 + \theta tr\ F \wedge F + cL^4 tr\ F^4 + \dots .$$

The coefficients c should be computable functions of $\tau = 4\pi/g_4^2 + i\theta$, and by the geometric origin of τ, should satisfy S-duality.

If we follow the chain $(2,0) \to D = 5 \to D = 4$, we can use these properties to try to constrain $D = 5$ MSYM. The perturbative argument is straightforward: compactify $D = 5$ MSYM on a circle of radius R_4, then

$$g_5^2 \int dp_4 \; f(p_4) \to \frac{R_5}{R_4} \sum_n f\left(\frac{n}{R_4}\right)$$

and the perturbative expansion is a series in $g_4^2 = R_5/R_4$.

Furthermore, the limit $L \to 0$ takes $R_4, R_5 \propto L \to 0$, so the 5d KK states go to infinite energy and can be dropped. This is the usual argument for the relation between $(2,0)$ theory, $D = 5$ and $D = 4$ MSYM.

$$g_5^2 \int dp_4 \; f(p_4) \to \frac{R_5}{R_4} \sum_n f\left(\frac{n}{R_4}\right)$$

However, UV divergences in $D = 5$ can potentially spoil this argument. These clearly come from states with nonzero p_4 (since the $D = 4$ theory was finite) and start to appear at energies $E \sim 1/R_4$. Such KK modes look like $D = 4$ particles with mass $M \sim n/R_4$.

Because the underlying theory is $(2,0)$ theory, there are no actual UV divergences – an apparent $D = 5$ UV divergence, is cutoff at the scale $\Lambda \sim 1/R_5$. Compared to pure $D = 4$ MSYM, one gets finite quantum corrections from states with $1/R_4 \leq E \leq 1/R_5$.

Thus, an apparent $D = 5$ divergence Λ^n, produces a correction

$$\left(\frac{\Lambda}{M}\right)^n \sim \left(\frac{R_4}{R_5}\right)^n \sim \tau^n,$$

or $\log \tau$ for a log divergence. Note that L has cancelled out, and these corrections need not disappear as $L \to 0$.

$$g_5^2 \int dp_4 \; f(p_4) \to \frac{R_5}{R_4} \sum_n f\left(\frac{n}{R_4}\right)$$

However, UV divergences in $D = 5$ can potentially spoil this argument. These clearly come from states with nonzero p_4 (since the $D = 4$ theory was finite) and start to appear at energies $E \sim 1/R_4$. Such KK modes look like $D = 4$ particles with mass $M \sim n/R_4$.

Because the underlying theory is $(2, 0)$ theory, there are no actual UV divergences – an apparent $D = 5$ UV divergence, is cutoff at the scale $\Lambda \sim 1/R_5$. Compared to pure $D = 4$ MSYM, one gets finite quantum corrections from states with $1/R_4 \leq E \leq 1/R_5$.

Thus, an apparent $D = 5$ divergence Λ^n, produces a correction

$$\left(\frac{\Lambda}{M}\right)^n \sim \left(\frac{R_4}{R_5}\right)^n \sim \tau^n,$$

or $\log \tau$ for a log divergence. Note that L has cancelled out, and these corrections need not disappear as $L \to 0$.

$$g_5^2 \int dp_4 \; f(p_4) \to \frac{R_5}{R_4} \sum_n f\left(\frac{n}{R_4}\right)$$

However, UV divergences in $D = 5$ can potentially spoil this argument. These clearly come from states with nonzero p_4 (since the $D = 4$ theory was finite) and start to appear at energies $E \sim 1/R_4$. Such KK modes look like $D = 4$ particles with mass $M \sim n/R_4$.

Because the underlying theory is $(2, 0)$ theory, there are no actual UV divergences – an apparent $D = 5$ UV divergence, is cutoff at the scale $\Lambda \sim 1/R_5$. Compared to pure $D = 4$ MSYM, one gets finite quantum corrections from states with $1/R_4 \leq E \leq 1/R_5$.

Thus, an apparent $D = 5$ divergence Λ^n, produces a correction

$$\left(\frac{\Lambda}{M}\right)^n \sim \left(\frac{R_4}{R_5}\right)^n \sim \tau^n,$$

or $\log \tau$ for a log divergence. Note that L has cancelled out, and these corrections need not disappear as $L \to 0$.

Taken at face value, this argument suggests that $D = 5$ MSYM must be *UV finite*, to prevent these problematic corrections. However this would be too quick:

- There might be other positive powers of L in front of these corrections, and
- There is a problem with taking $g_4 \to 0$ at fixed L.

To explain the second point, since

$$R_4 = \frac{L}{g_4}, \qquad R_5 = g_4 L,$$

taking $g_4 \to 0$ at fixed L decompactifies the fifth dimension. We need to take $L \to 0$ faster than $g_4 \to 0$ to keep the four dimensional interpretation.

Thus, compactified $D = 5 \to D = 4$ results are only guaranteed to have a clear interpretation, if we can compute them at finite g_4. There is one case where we can unambiguously do this, namely the $tr\ F^4$ term, which receives no higher loop corrections. Summing the one-loop contributions of KK modes with $p_4 \neq 0$, we find a coefficient

$$\sum_{p_4 \neq 0} \frac{g_4^4}{p_4^4} = \zeta(4) L^4$$

since the g_4 dependence of p_4 cancels that of the numerator.

This result is not S-dual. It can be promoted to an S-dual result by the ansatz of summing over both p_4 and p_5, leading to the coefficient

$$\sum_{(m,n) \neq (0,0)} \left(\frac{(Im\,\tau)^2}{|m\tau + n|^2} \right)^2 = \zeta(4) E(\tau, 2) L^4,$$

where $E(\tau, 2)$ is a nonholomorphic Eisenstein series. As a protected amplitude, this should be checkable in string theory, perhaps IIa little string theory (the throat region of NS 5-branes) on T^2.

Thus, compactified $D = 5 \to D = 4$ results are only guaranteed to have a clear interpretation, if we can compute them at finite g_4. There is one case where we can unambiguously do this, namely the $tr\ F^4$ term, which receives no higher loop corrections. Summing the one-loop contributions of KK modes with $p_4 \neq 0$, we find a coefficient

$$\sum_{p_4 \neq 0} \frac{g_4^4}{p_4^4} = \zeta(4) L^4$$

since the g_4 dependence of p_4 cancels that of the numerator.

This result is not S-dual. It can be promoted to an S-dual result by the ansatz of summing over both p_4 and p_5, leading to the coefficient

$$\sum_{(m,n) \neq (0,0)} \left(\frac{(Im\,\tau)^2}{|m\tau + n|^2} \right)^2 = \zeta(4) E(\tau, 2) L^4,$$

where $E(\tau, 2)$ is a nonholomorphic Eisenstein series. As a protected amplitude, this should be checkable in string theory, perhaps IIa little string theory (the throat region of NS 5-branes) on T^2.

As we will explain in detail later, the leading candidate divergent amplitude in $D = 5$ MSYM, at six loops, is indeed log divergent, leading to a term in the 4d effective action

$$c \cdot L^6 \tau^4 tr\ D^2 F^4 \times \log \tau \tag{2}$$

where $\log \tau = \log(\Lambda / M_4)$, the τ^4 comes from g_4^{14} for a four-point six loop amplitude and $1/g_4^6$ relating $1/M_5^6$ to L^6. This makes sense at fixed g_4 and looks like an example of the "non-regular" UV contribution we discussed above.

Of course, there is every reason to expect that higher loop computations will lead to higher and higher power divergences. By dimensional analysis, all contributions to $D^2 F^4$ go as L^6. What is less obvious, but can be checked, is that the leading divergences all come with the same power of τ. This is because the loop counting g_4^2 is compensated by the cutoff relation $\Lambda / M_4 = 1/g_4^2$. Thus there is no loop counting parameter.

Although it might be that this perturbation theory is simply nonsense, if we are summing over the correct states, this would be (to my mind) a very strange way for it to break down. In every other QFT, either there are new degrees of freedom, or the series makes some sort of qualitative sense (as in critical phenomena, QCD, semiclassical effects, etc.). Thus it is worth considering the alternative, that we have left out states which cancel the UV divergences. By the previous discussion, these **must be** the BPS states of our previous discussion, the self-dual "particles".

Thus we come to the hypothesis of 1012.2880 and of Lambert, Papageorgakis and Schmidt-Sommerfeld 1012.2882, that $D = 5$ MSYM is more than a low-energy effective description of $(2, 0)$ theory – it already contains all the degrees of freedom of $(2, 0)$ theory, we just need to find out how to incorporate them in our computations.

Although it might be that this perturbation theory is simply nonsense, if we are summing over the correct states, this would be (to my mind) a very strange way for it to break down. In every other QFT, either there are new degrees of freedom, or the series makes some sort of qualitative sense (as in critical phenomena, QCD, semiclassical effects, etc.). Thus it is worth considering the alternative, that we have left out states which cancel the UV divergences. By the previous discussion, these **must be** the BPS states of our previous discussion, the self-dual "particles".

Thus we come to the hypothesis of 1012.2880 and of Lambert, Papageorgakis and Schmidt-Sommerfeld 1012.2882, that $D = 5$ MSYM is more than a low-energy effective description of $(2, 0)$ theory – it already contains all the degrees of freedom of $(2, 0)$ theory, we just need to find out how to incorporate them in our computations.

The most straightforward way to work on this hypothesis is to define 5d MSYM with another auxiliary cutoff, take it to infinity and see what happens. This is not totally ridiculous as 5d MSYM is not chiral and could be put on the lattice, along the lines of Catterall hep-lat/0503036. Another approach is "deconstruction" as in Arkani-Hamed, Cohen and Georgi 0104005, in which we discretize the 5th dimension as the nodes and links of an $\mathcal{N} = 2$ 4d $\hat{A}_k$ quiver gauge theory (*i.e.*, a discretized circle). Giving a vev $v \cdot \mathbf{1}$ to each link matrix couples adjacent nodes with a discretized derivative term $v^2 tr\ (A_i - A_{i+1})^2$. This has the advantage of keeping 8 supercharges manifest.

It was reconsidered in Lambert *et al* 1212.3337 who rederive the old claim of Arkani-Hamed *et al* 0110146 that as $k \to \infty$ this becomes the $(2,0)$ theory, with the 6d KK modes visible as the S-dual partners of the W bosons with momentum in the 5th dimension.

These straightforward approaches give meaning to (and perhaps imply) the hypothesis, but do not directly answer questions such as how UV divergences in 5d are cutoff, or how an interacting 6d CFT can exist at all.

The most straightforward way to work on this hypothesis is to define 5d MSYM with another auxiliary cutoff, take it to infinity and see what happens. This is not totally ridiculous as 5d MSYM is not chiral and could be put on the lattice, along the lines of Catterall hep-lat/0503036.
Another approach is "deconstruction" as in Arkani-Hamed, Cohen and Georgi 0104005, in which we discretize the 5th dimension as the nodes and links of an $\mathcal{N} = 2$ 4d $\hat{A}_k$ quiver gauge theory (*i.e.*, a discretized circle). Giving a vev $v \cdot \mathbf{1}$ to each link matrix couples adjacent nodes with a discretized derivative term $v^2 tr\ (A_i - A_{i+1})^2$.
This has the advantage of keeping 8 supercharges manifest.
It was reconsidered in Lambert *et al* 1212.3337 who rederive the old claim of Arkani-Hamed *et al* 0110146 that as $k \to \infty$ this becomes the $(2, 0)$ theory, with the 6d KK modes visible as the S-dual partners of the W bosons with momentum in the 5th dimension.
These straightforward approaches give meaning to (and perhaps imply) the hypothesis, but do not directly answer questions such as how UV divergences in 5d are cutoff, or how an interacting 6d CFT can exist at all.

An obstacle to understanding 5d and 6d microscopic definitions from deconstruction is that in the 5d continuum limit, the node 4d gauge couplings G_4^2 go to infinity. This is because the action for the 5d zero modes is

$$S = \frac{1}{G_4^2} \sum_{i=1}^{k} tr\ F_i^2 + \dots \tag{3}$$

$$\to \frac{k}{G_4^2} tr\ F^2 + \dots \tag{4}$$

and the resulting YM coupling (in 4d or 5d) is proportional to G_4^2/k. Thus, the existence of the nodes does not seem helpful. One needs to go to 5d momentum space to work with the theory, and back to our previous description. But this does give another argument that the naive cutoff prescription is reasonable:

$$\int_{-\Lambda}^{\Lambda} dp_5 \to \sum_{n_5 > -k/2}^{k/2} \tag{5}$$

This brings us back to the idea that extra states in 5d MSYM, presumably the instantonic particles, provide the UV cutoff. Is it reasonable to look for this in perturbation theory? One might object that solitons generally have a form factor and their loop contributions will be exponentially suppressed.

This objection has been answered by Papageorgakis and Roysten (talk at String-Math and to appear). Using crossing symmetry, the soliton pair creation amplitude can be related to a form factor and computed in the Manton approximation:

$$\langle P'|J_\mu(x)|P\rangle = \int d^n m\, \Psi^*(m) J_\mu(\phi(m))\Psi(m), \qquad (6)$$

where m are coordinates on soliton moduli space, $\phi(m)$ is the corresponding field configuration, and $\Psi(m)$ is the wave function for a particular soliton state.

This brings us back to the idea that extra states in 5d MSYM, presumably the instantonic particles, provide the UV cutoff. Is it reasonable to look for this in perturbation theory? One might object that solitons generally have a form factor and their loop contributions will be exponentially suppressed.
This objection has been answered by Papageorgakis and Roysten (talk at String-Math and to appear). Using crossing symmetry, the soliton pair creation amplitude can be related to a form factor and computed in the Manton approximation:

$$\langle P'|J_\mu(x)|P\rangle = \int d^n m\, \Psi^*(m) J_\mu(\phi(m))\Psi(m), \tag{6}$$

where m are coordinates on soliton moduli space, $\phi(m)$ is the corresponding field configuration, and $\Psi(m)$ is the wave function for a particular soliton state.

One finds that, as expected, a soliton of size R has a form factor $\exp -R|p - p'|$ which is exponentially suppressed at large momentum transfer.

However, in the case at hand of self-dual solutions, the moduli include a scale size which can be arbitrarily small. After integrating over the scale size, the form factor becomes powerlike. Thus the instantonic particles could contribute at finite order in the coupling.

In principle we could use these techniques to constrain or compute their couplings. We might then work with the ansatz we used at one loop, and extend the sums over KK momenta p_4 to double sums over (p_4, p_5). Thus we get an L-loop amplitude as a sum over finite 4d amplitudes, similar to MSYM amplitudes, with massive internal states:

$$A_L \equiv \prod_{i=1}^{L} \sum_{p_4^i, p_5^i \in \mathbb{Z}} \int d^{4L}p \text{ integrand}. \tag{7}$$

One finds that, as expected, a soliton of size R has a form factor $\exp -R|p - p'|$ which is exponentially suppressed at large momentum transfer.

However, in the case at hand of self-dual solutions, the moduli include a scale size which can be arbitrarily small. After integrating over the scale size, the form factor becomes powerlike. Thus the instantonic particles could contribute at finite order in the coupling.

In principle we could use these techniques to constrain or compute their couplings. We might then work with the ansatz we used at one loop, and extend the sums over KK momenta p_4 to double sums over (p_4, p_5). Thus we get an L-loop amplitude as a sum over finite 4d amplitudes, similar to MSYM amplitudes, with massive internal states:

$$A_L \equiv \prod_{i=1}^{L} \sum_{p_4^i, p_5^i \in \mathbb{Z}} \int d^{4L}p \text{ integrand.} \qquad (7)$$

One finds that, as expected, a soliton of size R has a form factor $\exp -R|p-p'|$ which is exponentially suppressed at large momentum transfer.
However, in the case at hand of self-dual solutions, the moduli include a scale size which can be arbitrarily small. After integrating over the scale size, the form factor becomes powerlike. Thus the instantonic particles could contribute at finite order in the coupling.
In principle we could use these techniques to constrain or compute their couplings. We might then work with the ansatz we used at one loop, and extend the sums over KK momenta p_4 to double sums over (p_4, p_5). Thus we get an L-loop amplitude as a sum over finite 4d amplitudes, similar to MSYM amplitudes, with massive internal states:

$$A_L \equiv \prod_{i=1}^{L} \sum_{p_4^i, p_5^i \in \mathbb{Z}} \int d^{4L}p \text{ integrand}. \tag{7}$$

We can also try to guess the couplings. One constraint is S-duality in 4d. For example, the ansatz which led to the one-loop result we discussed earlier constrains the KK mode couplings to the $D = 5$ MSYM massless states, as was worked out in Czech, Huang and Rozali 1110.2791. In a generic gauge theory, there are many one-loop four point diagrams; only in special cases such as MSYM do these reduce to the box diagram with no bubbles or triangles. Combining this constraint with the lack of KK momentum dependence in the one-loop amplitude leads to a vertex

$$f_{ijk} A^i_\mu B(p_5)^j_{\nu\rho} \partial^{[\mu} B(-p_5)^{k,\nu\rho]} \tag{8}$$

where A_μ is the 5d gauge field and B are the massive KK tensors, plus a generalization with one more parameter.
This is a sort of massive chiral generalization of the 5d YM vertex.

Another, probably stronger constraint is that if we are really going to get finite amplitudes to all orders, this will almost certainly be because of a supersymmetry involving the new modes. This may sound problematic as it is a short step to saying that we are computing with a 6d action containing the new states, but there is no conventional local action describing nonabelian tensors in 6d.

Actually, there has been some progress in writing actions in 5d which contain the entire spectrum of BPS Kaluza-Klein modes, by Ho, Huang and Matsuo 1104.4040; Samtleben, Sezgin and Wimmer, 1108.4060; and Bonetti, Grimm and Hohenegger 1209.3017. These contain the vertex derived by Czech et al.

The main loophole they exploit is that these actions are nonlocal in the sixth dimension – at the very least they treat the zero modes differently from the nonzero modes, and some work uses explicit inverse derivatives. So, these would not be acceptable fundamental formulations – but they might make sense as descriptions of couplings to 5d solitons. It will be interesting to try to understand the UV finiteness of the theory in these terms.

Another, probably stronger constraint is that if we are really going to get finite amplitudes to all orders, this will almost certainly be because of a supersymmetry involving the new modes. This may sound problematic as it is a short step to saying that we are computing with a 6d action containing the new states, but there is no conventional local action describing nonabelian tensors in 6d.

Actually, there has been some progress in writing actions in 5d which contain the entire spectrum of BPS Kaluza-Klein modes, by Ho, Huang and Matsuo 1104.4040; Samtleben, Sezgin and Wimmer, 1108.4060; and Bonetti, Grimm and Hohenegger 1209.3017. These contain the vertex derived by Czech et al.

The main loophole they exploit is that these actions are nonlocal in the sixth dimension – at the very least they treat the zero modes differently from the nonzero modes, and some work uses explicit inverse derivatives. So, these would not be acceptable fundamental formulations – but they might make sense as descriptions of couplings to 5d solitons. It will be interesting to try to understand the UV finiteness of the theory in these terms.

Even if it worked, the ansatz of letting instantonic particles run around loops would raise many questions. The basic point that we have not understood is that, because of their scale size modulus, the instantonic particles are really extended objects. I repeat the point that this is an inevitable part of a nontrivial CFT in $D > 4$.

Time evolution of a single instantonic particle consists not just of motion of its center of mass, but also of its scale size. Should we think of the scale size as another asymptotic direction, something like the radial dimension of AdS/CFT ?

The charge k-instanton moduli space has real dimension $4kN$. It has limits which look like k $SU(2)$ instantons, but the general configuration does not. There have been many suggestions that an $SU(N)$ instanton can "fractionate" into N constituents. This idea was explored in $D = 5$ MSYM by Collie and Tong 0905.2267, who develop an interesting analogy to solitons in the $D = 3$ $\mathbb{CP}^N$ model.

Even if it worked, the ansatz of letting instantonic particles run around loops would raise many questions. The basic point that we have not understood is that, because of their scale size modulus, the instantonic particles are really extended objects. I repeat the point that this is an inevitable part of a nontrivial CFT in $D > 4$.

Time evolution of a single instantonic particle consists not just of motion of its center of mass, but also of its scale size. Should we think of the scale size as another asymptotic direction, something like the radial dimension of AdS/CFT ?

The charge k-instanton moduli space has real dimension $4kN$. It has limits which look like k $SU(2)$ instantons, but the general configuration does not. There have been many suggestions that an $SU(N)$ instanton can "fractionate" into N constituents. This idea was explored in $D = 5$ MSYM by Collie and Tong 0905.2267, who develop an interesting analogy to solitons in the $D = 3$ $\mathbb{CP}^N$ model.

$D = 5$ MSYM at six loops

Given the problematic interpretation of the $D = 5$ six loop divergence, and the analogy to other interesting problems such as the possible seven loop divergence in $N = 8$ supergravity, Zvi Bern, John Joseph Carrasco, Lance Dixon, Matt von Hippel and Henrik Johansson and I were motivated to actually do the computation.

Although a six loop computation may seem prohibitively difficult, a combination of recent advances in techniques for working with gauge theory amplitudes, and advances in numerical integration of Feynman diagrams, made it possible. The basic formalism for MSYM amplitudes is largely independent of dimension, so $D = 4$, $N = 4$ SYM results can be adapted to this case.

To get the contribution to D^2F^4, we need to insert four external gluons and take the limit of zero external momenta. Using superspace arguments this can be reduced to a vacuum diagram, but with some doubled propagators. In general there are also complicated numerator factors, but these can be greatly simplified by working out the on-shell amplitude with the method of unitarity cuts. The resulting diagrams have no UV subdivergences in $D = 5$.

Further simplifications are obtained by working with on-shell $D = 6$ superspace, as in arXiv:1010.0494 (Bern, Carrasco, Dennen, Huang, and Ita).

Finally, one can use integration by parts to derive many relations between different topologies. For the present amplitude, *all* of the topologies can be re-expressed in terms of the pentagonal prism with various numerator factors and doubled propagators.

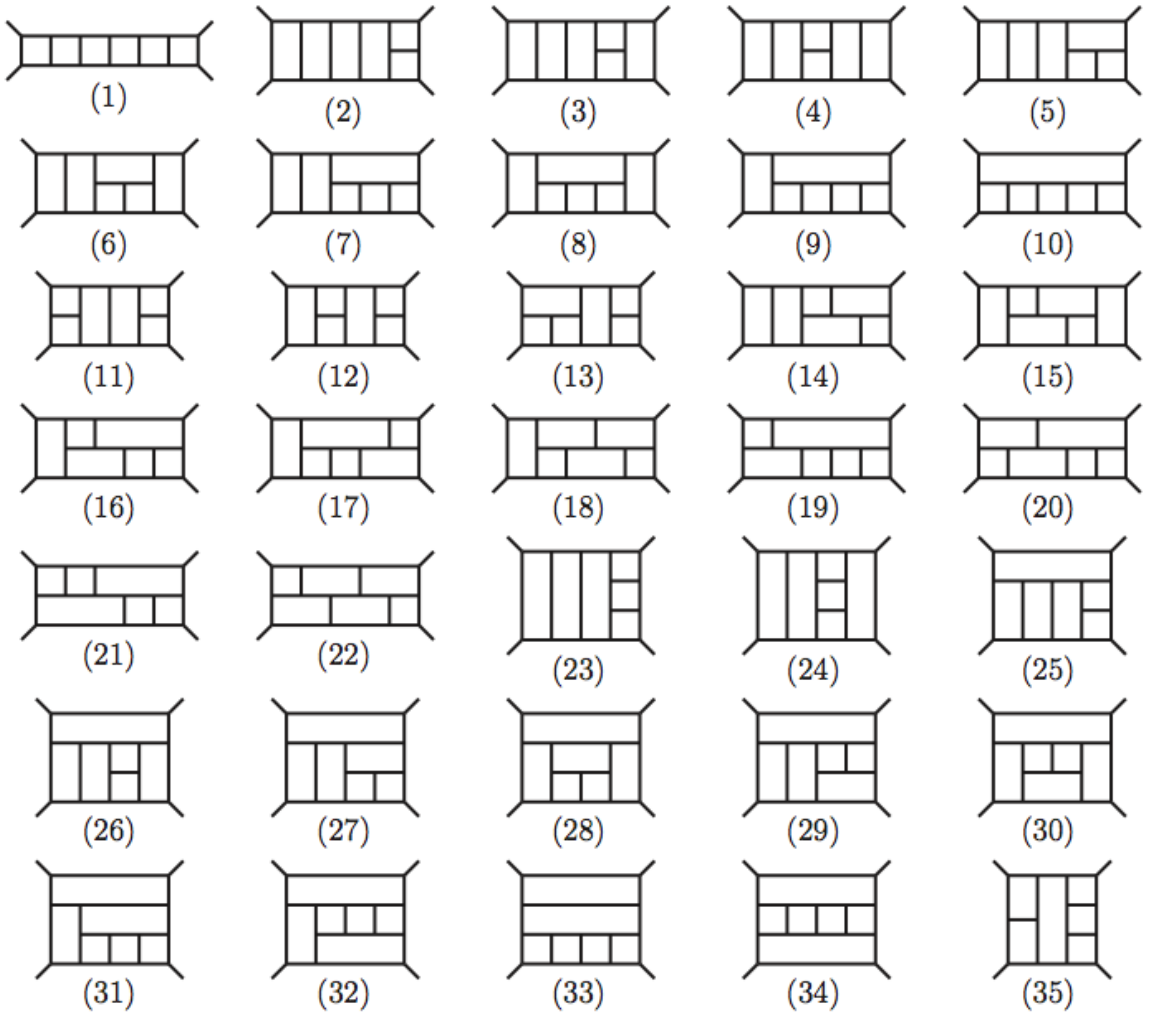
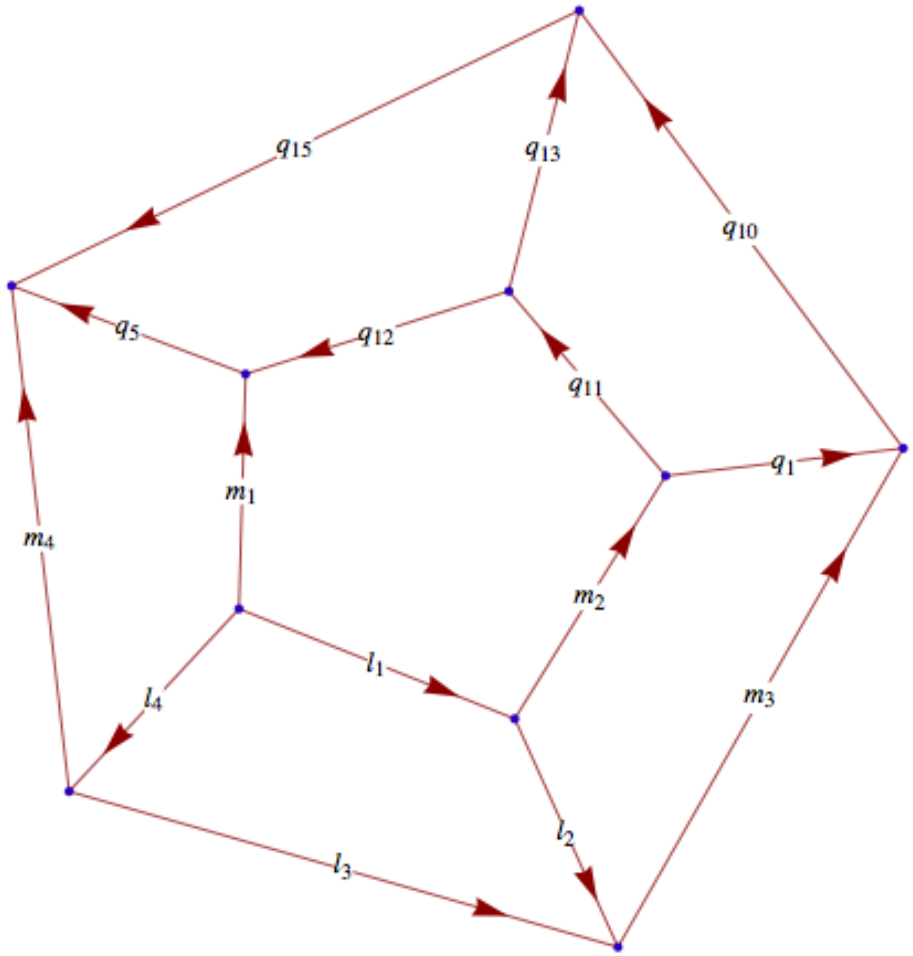


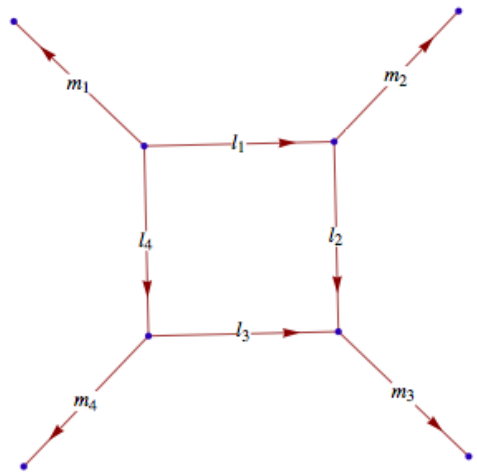
FIG. 1: Graphs 1 through 35 for the planar six-loop four-point amplitude.

There are 68 different triangle-free six loop planar graphs which contribute; here are 35 of them.

Using integration by parts identities, these vacuum diagrams can all be reduced to the pentagonal prism:



It turns out that all of the numerator and denominator factors can be placed on a single one-loop subdiagram of the prism,



leading to an explicit integrand, with the dressing factor

$$F = \frac{s^2}{l_1^2 l_3^2} + \frac{s\,t}{l_2^2 l_4^2} + 2\frac{s\,t}{m_1^2 m_2^2} + \frac{s\,l_2^2}{m_3^2 l_3^2} + \frac{s\,l_4^2}{m_4^2 l_3^2} + \frac{s\,l_1^2}{m_2^2 l_2^2} + \frac{s\,l_1^2}{m_1^2 l_4^2} + \frac{s\,l_3^2}{m_3^2 l_2^2} + \frac{s\,l_3^2}{m_4^2 l_4^2} - 2\frac{l_1^2}{l_3^2}$$

where $s = (m_1 + m_4)^2$ and $t = (m_1 + m_2)^2$.

It turns out that by further use of IBP relations, one can simplify this to

$$\begin{aligned} F' &= 5\left(\frac{(l_1+l_3)^2}{l_2^2}+\frac{l_3^2}{l_1^2}+\frac{l_1^2}{l_3^2}\right)-1 \\ &= 5\frac{(l_1+l_3)^2}{l_2^2}+\frac{9}{2}\left(\frac{l_3^2}{l_1^2}+\frac{l_1^2}{l_3^2}\right)+\frac{1}{2}\frac{(l_1^2-l_3^2)^2}{l_1^2 l_3^2} \\ &> 0, \end{aligned}$$

so the D^2F^4 contribution is indeed log divergent.

We did not have these IBP relations until late in our work and thus we went on to numerically estimate the integral by Monte Carlo integration. The fifteen propagators of the prism mean that this will be a fifteen dimensional integral, which one might think would be easy for a computer to estimate.

However, the reason Monte Carlo is not used more often for Feynman integrals, is because it behaves badly for divergent integrands. Although in the present case the integrand has only integrable divergences, this is bad enough – the error estimate is proportional to the variance of the integrand, but this is not square integrable.

We did not have these IBP relations until late in our work and thus we went on to numerically estimate the integral by Monte Carlo integration. The fifteen propagators of the prism mean that this will be a fifteen dimensional integral, which one might think would be easy for a computer to estimate.

However, the reason Monte Carlo is not used more often for Feynman integrals, is because it behaves badly for divergent integrands. Although in the present case the integrand has only integrable divergences, this is bad enough – the error estimate is proportional to the variance of the integrand, but this is not square integrable.

Now given any integrable endpoint divergence, for example

$$\int \frac{dx_1\, dx_2\, dx_3}{x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}},$$

it can be resolved by the change of variables $y_i = x_i^{1-\alpha_i}$, and then done by Monte Carlo.
The problem is that the integrand has many endpoints – up to 15! in this case – and each one requires a different change of variable.

An approach to deal with this is **sector decomposition**, as developed by Binoth and Heinrich (hep-ph/0004013, 0402265), Bogner and Weinzierl (arXiv:0709.4092, 0806.4307), and A. and V. Smirnov and M. Tentyukov (arXiv:0912.0158). The idea is simply to decompose the integration region (the unit simplex) into many sectors, each containing at most one singularity, and then integrate each sector separately. This can also be used for UV and IR divergent integrands, by making subtractions or using Mellin-Barnes techniques.

Now given any integrable endpoint divergence, for example

$$\int \frac{dx_1\, dx_2\, dx_3}{x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}},$$

it can be resolved by the change of variables $y_i = x_i^{1-\alpha_i}$, and then done by Monte Carlo.
The problem is that the integrand has many endpoints – up to 15! in this case – and each one requires a different change of variable.

An approach to deal with this is **sector decomposition**, as developed by Binoth and Heinrich (hep-ph/0004013, 0402265), Bogner and Weinzierl (arXiv:0709.4092, 0806.4307), and A. and V. Smirnov and M. Tentyukov (arXiv:0912.0158). The idea is simply to decompose the integration region (the unit simplex) into many sectors, each containing at most one singularity, and then integrate each sector separately. This can also be used for UV and IR divergent integrands, by making subtractions or using Mellin-Barnes techniques.

The basic problem can be seen by looking at a planar scalar field theory Feynman diagram with E edges, L loops, and no numerator factors:

$$I = \int \prod_{a=1}^{L} d^D p_a \prod_{i=1}^{E} \frac{1}{p_i^2 + m_i^2},$$

where the edge momenta p_i depend linearly on the loop momenta p_a in a way determined by the graph topology Γ.

In this case, the Feynman parameterization is

$$I = C \int_{\alpha_i \geq 0} \prod_{i=1}^{E} d\alpha_i \left(\det{}' \Delta_\Gamma(\alpha)\right)^{-D/2},$$

where $\Delta_\Gamma(\alpha)$ is a graph Laplacian which acts on functions on the vertices of the dual graph (*i.e.* faces of Γ) as

$$\Delta_\Gamma(\alpha) = d^\dagger(\alpha) \cdot d(\alpha); \qquad d(\alpha) f = \sum_{i=1}^{E} \frac{e_i}{\alpha_i} \left(f(\mathrm{start}(i)) - f(\mathrm{end}(i))\right).$$

This determinant is equal to the Kirchhoff polynomial of Γ, which is a sum over spanning trees $T \subset \Gamma$,

$$\det{}' \Delta_\Gamma(\alpha_i) = \sum_T \prod_{i \in T} \alpha_i. \tag{9}$$

Thus one has an explicit integrand, which diverges wherever this polynomial vanishes.

The Kirchhoff polynomial vanishes whenever the graph Laplacian obtains extra zero modes. It can be shown that, for $\alpha_i \geq 0$, this only happens when the graph obtained by removing all edges with $\alpha_i = 0$, is disconnected. There are many ways to do this and thus the vanishing (or discriminant) locus is a complicated subvariety of the unit simplex, with components of various dimensions. The goal of sector decomposition is to subdivide the simplex into sectors each intersecting at most one component of the discriminant locus.

This determinant is equal to the Kirchhoff polynomial of Γ, which is a sum over spanning trees $T \subset \Gamma$,

$$\det{}' \Delta_\Gamma(\alpha_i) = \sum_T \prod_{i \in T} \alpha_i. \tag{9}$$

Thus one has an explicit integrand, which diverges wherever this polynomial vanishes.

The Kirchhoff polynomial vanishes whenever the graph Laplacian obtains extra zero modes. It can be shown that, for $\alpha_i \geq 0$, this only happens when the graph obtained by removing all edges with $\alpha_i = 0$, is disconnected. There are many ways to do this and thus the vanishing (or discriminant) locus is a complicated subvariety of the unit simplex, with components of various dimensions. The goal of sector decomposition is to subdivide the simplex into sectors each intersecting at most one component of the discriminant locus.

The problem of separating out the various ways in which the Kirchoff polynomial can vanish is a problem in resolution of singularities, and can be treated by these methods. The basic operation is a blow-up: we choose a variable α_i and make the change of variables

$$x_i = \alpha_i; \qquad x_i x_j = \alpha_j \ \forall j \neq i.$$

This converts the region of integration $0 \leq \alpha_j \leq \alpha_i \ \forall j \neq i$ into the region $0 \leq x_j \leq 1 \ \forall j \neq i$. Thus, the original N-dimensional region $\alpha_i \geq 0 \ \forall i$ turns into a union of N "primary sectors," each with an integrand homogeneous in x_i. The x_i integral is easy to do (and produces the log divergence in our case) leaving a nontrivial integral over the $N-1$-dimensional hypercube for each primary sector.

Having factored out one variable x_i, the rest of the denominator will take the form

$$\mathcal{D} = \sum_T \prod_{j \in T, j \neq i} x_j$$

and is no longer homogeneous – some terms are degree L and others are degree $L-1$.

Now, in each primary sector, we iterate the blow-up process, choosing a subset of variables x_k with $k \in S$, and dividing the sector into $|S|$ new sectors labelled by $k \in S$. In sector k, we make the change of variables

$$x'_k = x_k; \qquad x'_k x'_j = x_j \; \forall j \neq k.$$

This subdivides the $N-1$-cube into $|S|$ unit cubes, in each of which the denominator will have a higher power of some variable x'_k. If the resulting denominator has the form

$$\mathcal{D} = x_{k_1}^{n_1} \dots x_{k_r}^{n_r} \left(1 + x \dots\right),$$

then the singularity has been resolved in that sector, and the integral can be done by change of variable (if the $n_i > -1$) or other procedures.

This iterative procedure can be related to resolution of a singularity by blow-up, and Hironaka's proof of resolution of singularities can be adapted to show that there always exists a decomposition into nonsingular sectors and that it can be found by an algorithm.

Several such algorithms have been implemented in computer packages. We adapted A. and V. Smirnov's FIESTA 2 package, which uses Mathematica for symbolic manipulation, then sends the integrands for each sector to a C++ Monte Carlo integrator (Vegas). Such a sector integral at 10^{-3} precision takes a few seconds on one CPU.

Among the sector decomposition algorithms of FIESTA 2 is a "heuristic" algorithm which follows the procedure we described, and chooses the set S at each step to be the one which leads to factoring out a maximal degree $\sum n_i$ monomial. This algorithm is not guaranteed to terminate, but often works in practice.

Applied to our six loop diagram, the heuristic algorithm works, producing about 10^6 sectors (the precise number depends on the term in the dressing factor). This is much less than 15! and the resulting integral can be done on a 1000 node cluster in a few hours. The result is

$$A_{L=6} = -\frac{1}{\epsilon}\frac{stuA_{tree}}{(4\pi)^{15}} \times (68.68 \pm 0.17). \tag{10}$$

An amusing and rather mysterious fact is that the known leading log divergent amplitudes in dimensions $D = 4 + 6/\ell$ can be fit to a simple formula:

$$A_\ell = (-1)^{\ell-1}\frac{1}{\epsilon}\frac{stuA_{tree}}{(4\pi)^{2L-3}}\left(b + c^{\ell+a/\ell}\right) \tag{11}$$

with $a = 3.99$, $b = 1.74 \times 10^{-5}$ and $c = 9.77$.

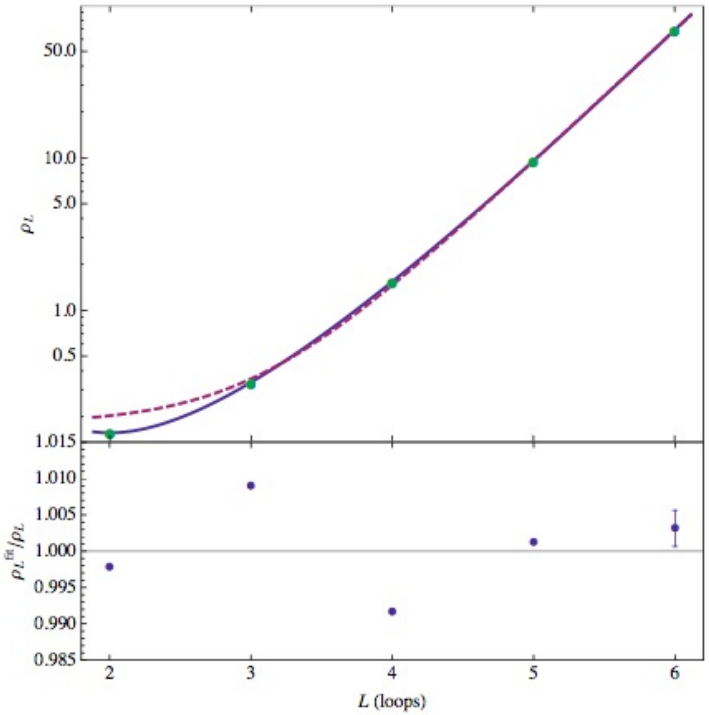
Applied to our six loop diagram, the heuristic algorithm works, producing about 10^6 sectors (the precise number depends on the term in the dressing factor). This is much less than 15! and the resulting integral can be done on a 1000 node cluster in a few hours. The result is

$$A_{L=6} = -\frac{1}{\epsilon}\frac{stuA_{tree}}{(4\pi)^{15}} \times (68.68 \pm 0.17). \tag{10}$$

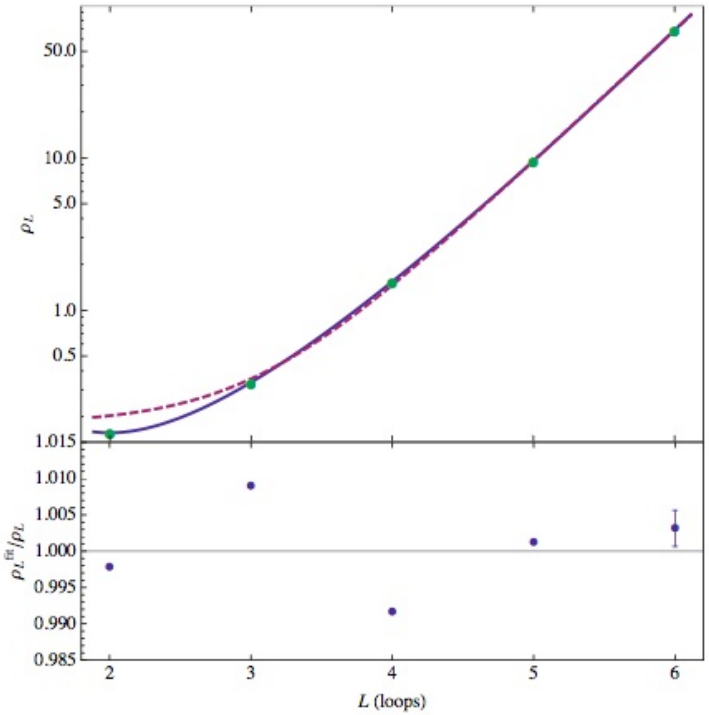
An amusing and rather mysterious fact is that the known leading log divergent amplitudes in dimensions $D = 4 + 6/\ell$ can be fit to a simple formula:

$$A_\ell = (-1)^{\ell-1}\frac{1}{\epsilon}\frac{stuA_{tree}}{(4\pi)^{2L-3}}\left(b + c^{\ell + a/\ell}\right) \tag{11}$$

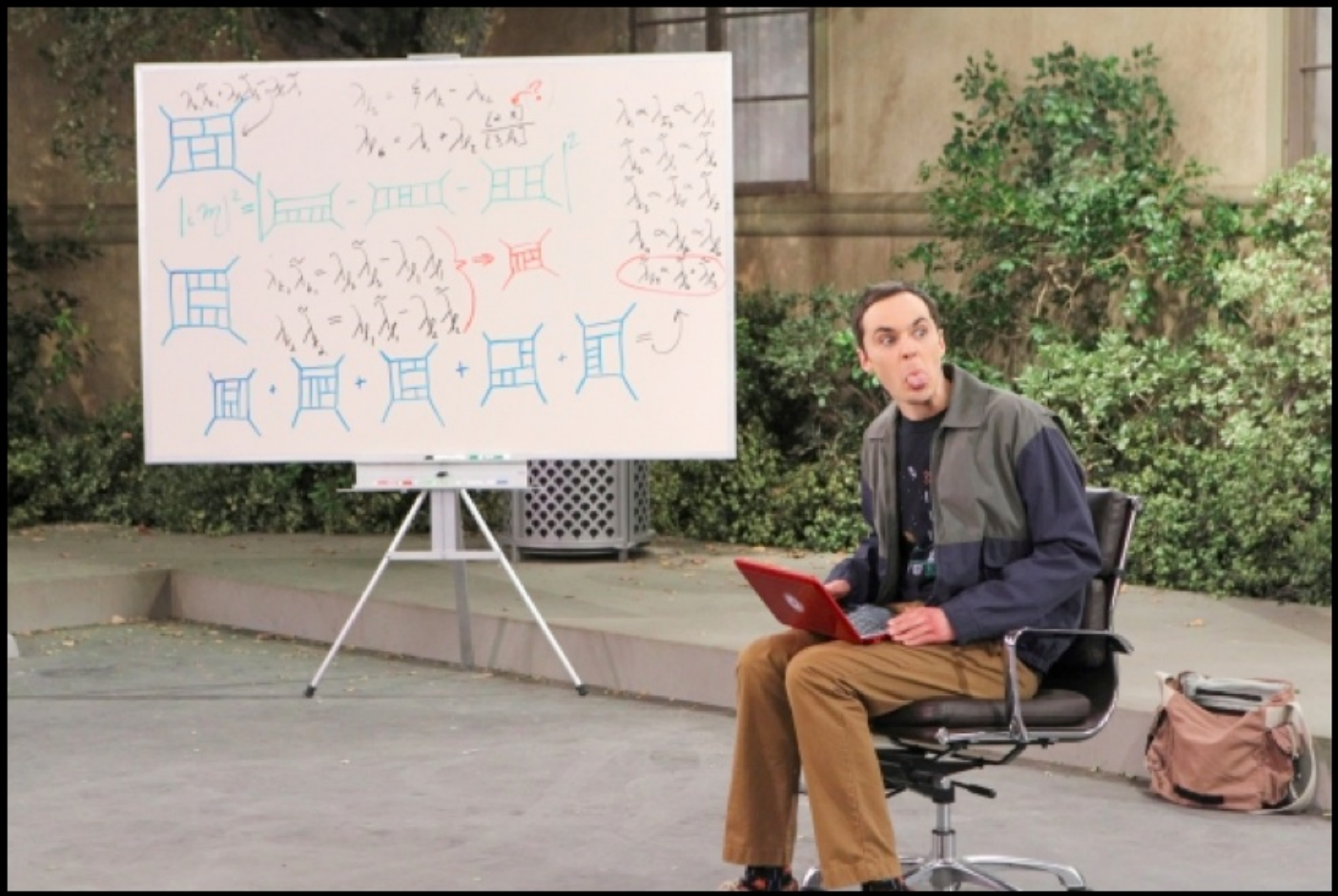
with $a = 3.99$, $b = 1.74 \times 10^{-5}$ and $c = 9.77$.



A seven loop calculation would be expected to lead to about 10^8 sectors and should also be doable, so if the leading potentially log divergent seven loop amplitude in $N = 8$ supergravity can be brought to a reasonable size, this long-standing question could also be settled by direct computation.



A seven loop calculation would be expected to lead to about 10^8 sectors and should also be doable, so if the leading potentially log divergent seven loop amplitude in $N = 8$ supergravity can be brought to a reasonable size, this long-standing question could also be settled by direct computation.



The competition in six loop computations.
From ``Big Bang Theory,'' CBS, Nov 29