New Physics, the Hierarchy Problem, and the String Theory Landscape

Notes of lecture course by A. Hebecker, 2018 and 2019

The plan is to give a concise but technical introduction to 'Physics Beyond the Standard Model' and early cosmology as seen from the perspective of string theory. In particular, the two hierarchy problems (of the cosmological constant and the electroweak scale) will be discussed in view of ideas like string theory landscape, eternal inflation and multiverse. The presentation will include critical points of view and alternative ideas and explanations. Basic knowledge of quantum field theory and general relativity (but not of string theory) will be assumed. Elements of string theory will be introduced as needed. Useful literature will also be mentioned as we go along. Two texts in particular stand out because they share the spirit of this course: The first is the set of lecture notes [1], which starts however at a much higher technical level. The other is the extensive textbook or even monograph [2]. Both are much more 'stringy' than the present notes, the emphasis being more on concrete string model constructions rather than on the physics of hierarchy problem and (multiverse) cosmology.

Contents

1 The Standard Model and its Hierarchy Problem(s) 3

1.1 Standard Model - the basic structure . . . 3

1.2 Standard Model - parameter count . . . 6

1.3 Effective field theories - cutoff perspective . . . 9

1.4 Effective field theories - $\text{QFT}_{\mathbf{UV}}$ vs. $\text{QFT}_{\mathbf{IR}}$ . . . 10

1.5 The Standard Model as an effective field theory . . . 12

1.6 The electroweak hierarchy problem . . . 15

1.7 Fine tuning . . . 17

1.8 Gravity and the cosmological constant problem . . . 20

2 Supersymmetry and Supergravity 23

2.1 SUSY algebra and superspace . . . 24

2.2 Superfields . . . 26

2.3 Chiral superfields . . . 27

2.4 SUSY-invariant lagrangians . . . 28

2.5 Wess-Zumino-type models . . . 29

2.6 Real Superfields . . . 30
2.7 SUSY breaking . . . 32
2.8 Supersymmetrizing the Standard Model . . . 33
2.9 Supersymmetric and SUSY breaking masses and non-renormalization . . . 35
2.10 The Minimal Supersymmetric Standard Model (MSSM) . . . 37
2.11 Supergravity - superspace approach . . . 39
2.12 Supergravity - component approach . . . 42

3 String Theory: Bosonic String **44**
3.1 Strings – basic ideas . . . 44
3.2 Symmetries, equations of motion, gauge choice . . . 47
3.3 Open string . . . 50
3.4 Quantization . . . 51
3.5 Explicit construction of physical states – open string . . . 56
3.6 Explicit construction of physical states – closed string . . . 58
3.7 The 26d action . . . 60

4 String Theory: Interactions and Superstring **63**
4.1 State-operator correspondence . . . 63
4.2 Scattering amplitudes . . . 64
4.3 World sheet supersymmetry . . . 66
4.4 World sheet supergravity . . . 68
4.5 Quantization of the superstring . . . 69
4.6 GSO or Gliozzi-Scherk-Olive projection . . . 72
4.7 Consistent type II superstring theories . . . 73
4.8 Other 10d theories . . . 76

5 10d actions and compactification **76**
5.1 10d supergravities and Type IIB as an example . . . 76
5.2 Kaluza-Klein compactification . . . 79
5.3 Towards Calabi-Yaus . . . 82
5.4 Homology and cohomogy . . . 86
5.5 Calabi-Yau moduli spaces . . . 91

5.6 Explicit parameterization of Calabi-Yau moduli spaces 94

6 The flux landscape 98

6.1 Compact geometries with p-form fluxes 99

6.2 Bousso-Polchinski model 102

6.3 The type-IIB flux landscape (GKP) 104

6.4 Kahler modulus stabilization and SUSY breaking (KKLT) 106

6.5 The anti-D3-brane uplift of KKLT 111

6.6 The Large Volume Scenario 115

6.7 Vacuum Statistics and the realizability fine tunings 115

7 Eternal Inflation and the Measure Problem 119

7.1 From slow-roll inflation to the eternal regime 119

7.2 Slow-roll inflation in the landscape 120

7.3 The measure problem 122

1 The Standard Model and its Hierarchy Problem(s)

This section assumes at least some familiarity with quantum field theory (QFT), including basics of regularization and renormalization. There exists a large number of excellent textbooks on this subject, such as [3–7]. The reader familiar with this topic will most probably also have some basic understanding of the Standard Model of Particle Physics, although this will not be strictly necesssary since we will introduce this so-called Standard Model momentarily. It is also treated at different levels of detail in most QFT texts, most notably in [4, 7]. Books devoted specifically to theoretical particle physics and the Standard Model include [8–11]. We will try to refer to some more specialized texts or even articles as we go along.

1.1 Standard Model - the basic structure

A possible definition of the Standard Model is as follows: It is the most general renormalizable field theory with gauge group

$$G_{SM} = SU(3) \times SU(2) \times U(1)\,, \tag{1.1}$$

three generations of fermions, and a scalar. These fields transform in the representations

$$(\mathbf{3},\mathbf{2})_{1/6} + (\bar{\mathbf{3}},\mathbf{1})_{-2/3} + (\bar{\mathbf{3}},\mathbf{1})_{1/3} + (\mathbf{1},\mathbf{2})_{-1/2} + (\mathbf{1},\mathbf{1})_{1} \qquad \text{and} \qquad (\mathbf{1},\mathbf{2})_{1/2} \tag{1.2}$$

respectively. Here the boldface numbers specify the representations of $SU(3)$ and $SU(2)$ via its dimension (in our case only singlet, fundamental or anti-fundamental occur, the latter denoted by an overline) and the index gives the $U(1)$ charge. The overall normalization of the latter is clearly convention dependent since there is no intrinsic way to normalize a $U(1)$ gauge field.[1]

If one adds gravity in its simplest and essentially unique theoretical formulation (Einstein's general relativity), then this data offers an almost complete fundamental description of the material world. This structural simplicity and the resulting small number of fundamental parameters (to be specified in a moment) is very remarkable. What is even more remarkable is the enormous underlying unification: So many very different macroscopic and microscopic phenomena which we observe in everyday life and in many natural sciences follow from such a (relatively) simple underlying theory.

Clearly, important caveats have already been noted above: The description is *almost* complete, the theory is *relatively* simple (not as simple as one would wish) and, maybe most importantly, it is only fundamental *to the extent that we can test it at the moment*. Quite possibly, more fundmental building blocks can be identified in the future. The rest of this course is about exactly these caveats and whether, based on those, theoretical progress is possible.

But first let us be more precise and explicit and turn the defining equations (1.1) and (1.2) into a field-theoretic lagrangian. Given the theoretically well-understood and experimentally tested rules of quantum field theory (QFT), this can be done unambiguously. The structure of the lagrangian is

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{matter} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}\,. \tag{1.3}$$

The gauge part is completely standard:

$$\mathcal{L}_{gauge} \;=\; -\frac{1}{4g_1^2}F^{(1)}_{\mu\nu}F^{(1)\,\mu\nu} \;-\; \frac{1}{2g_2^2}\,\mathrm{tr}\,F^{(2)}_{\mu\nu}F^{(2)\,\mu\nu} \;-\; \frac{1}{2g_3^2}\,\mathrm{tr}\,F^{(3)}_{\mu\nu}F^{(3)\,\mu\nu}\,. \tag{1.4}$$

Of course, one has to remember the conventional normalization $\mathrm{tr}(T_AT_B)=\delta_{AB}/2$ of the $SU(N)$ generators in the fundamental representation. The matter or, more precisely, the *fermionic* matter contribution reads

$$\mathcal{L}_{matter} = \sum_j \overline{\psi}_j i\not{D}_j\psi_j \qquad \text{with} \qquad (D_j)_\mu = \partial_\mu - iR_j(A_\mu) \tag{1.5}$$

with j running over left-handed quark doublets, right-handed up- and down-type quarks, lepton-doublet and right-handed leptons (each coming in three generations or families):

$$\psi_j \in \{\, \{q_L^a, (u_R^a)^c, (d_R^a)^c, l_L^a, (e_R^a)^c\},\; a=1,2,3\,\}\,. \tag{1.6}$$

The five types of fermions from q_L to e_R^c correspond precisely to the five terms in the direct sum in (1.2). Furthermore, $R_j(A_\mu)$ denotes the representation of $A_\mu \in Lie(G_{SM})$ appropriate for the fermion of type j. To make our conventions unambiguous, we have to specify in detail how we describe the spinor fields. One convenient choice (the one implicitly used above) is to always work with left-handed 4-component or Dirac spinors. In other words, we do *not* use general Dirac spinors built from Weyl spinors according to

$$\psi_D = \begin{pmatrix} \psi_\alpha \\ \overline{\chi}^{\dot\alpha} \end{pmatrix} \qquad \text{with} \qquad \alpha, \dot\alpha = 1,2\,. \tag{1.7}$$

[1] In our conventions the electric charge is given by $Q = T_3 + Y$.

Instead, all our 4-spinors are left-handed:

$$\psi = \begin{pmatrix} \psi_\alpha \\ 0 \end{pmatrix} . \tag{1.8}$$

In particular, this explains why we use the charge-conjugate of right-handed quarks and leptons as our fundamental fields, cf.

$$q_L = \begin{pmatrix} (q_L)_\alpha \\ 0 \end{pmatrix} \qquad \text{vs.} \qquad u_R^c = \begin{pmatrix} (u_R)_\alpha \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ (\overline{u}_R)^{\dot\alpha} \end{pmatrix}^c . \tag{1.9}$$

We see that the quantum numbers given in (1.2) can be viewed as referring to either these left-handed fields or to the corresponding 2-component Weyl spinors. The latter will in any case be very useful when talking about supersymmetry below.

The scalar or Higgs lagrangian is

$$\mathcal{L}_{Higgs} = -(D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi) \qquad \text{with} \qquad V(\Phi) = -m_H^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 , \tag{1.10}$$

where Φ is an $SU(2)$ doublet with charge 1/2 under $U(1)$ (the hypercharge-$U(1)$ or $U(1)_Y$). Finally, there are the Yukawa terms

$$\mathcal{L}_{Yukawa} = -\sum_{jk} \lambda_{jk}\overline{\psi}_j\psi_k^c\,\Phi \;+\; \text{h.c.} , \tag{1.11}$$

where the sum runs over all combinations of fields for which the relevant product of representations contains a gauge singlet. We left the group indices and their corresponding contraction implicit.

Note that, since all our fields are l.h. 4-component spinors, we have to write $\overline{\psi}\psi^c$ rather than simply $\overline{\psi}\psi$. The latter would be identically zero. Note also that the 4-spinor expression $\overline{\psi}\psi^c$ corresponds to $\overline{\psi}_{\dot\alpha}\overline{\psi}^{\dot\alpha}$ in terms of the Weyl spinor ψ_α contained in the 4-spinor ψ.

Crucially, the Higgs potential has a minimum with S^3 topology at $|\Phi| = v \simeq 174$ GeV, leading to spontaneous gauge symmetry breaking. One can choose the VEV to be real and aligned with the lower component of Φ, leading to the parameterization

$$\Phi = \begin{pmatrix} 0 \\ v + h/\sqrt{2} \end{pmatrix} . \tag{1.12}$$

It is easy to see that the symmetry breaking pattern is $SU(2)\times U(1)_Y \to U(1)_{em}$ (see problems). Three would-be Goldstone-bosons along the S^3 directions are 'eaten' by three of the four vector bosons of $SU(2)\times U(1)_Y$. This leads to the $W^\pm$ and Z bosons with masses $m_{W^\pm} \simeq 80$ GeV and $m_Z \simeq 90$ GeV. The surviving real Higgs scalar h is governed by

$$\mathcal{L} \supset -\frac{1}{2}(\partial h)^2 - \frac{1}{2}m_h^2 h^2 . \tag{1.13}$$

One can relate the parameters after symmetry breaking to those of the original lagrangian:

$$v^2 = m_H^2/(2\lambda) \qquad , \qquad m_h^2 = 4\lambda v^2 . \tag{1.14}$$

We have $m_h \simeq 125$ GeV, $m_H = m_h/\sqrt{2} = 88$ GeV, and $\lambda \simeq 0.13$.[2]

The surviving massless gauge boson is, of course, the familiar photon. Finally, one can check that the allowed Yukawa terms suffice to give all charged fermions a mass proportional to v. The three lightest quark masses are not directly visible to experiment since the confinement dynamics of the $SU(3)$ gauge theory (QCD) hides their effect. The three upper components of the lepton-doublets – the neutrinos – have $Q = 0$ and remain massless.

The reader should check explicitly which (three) Yukawa terms a allowed and that *no* further renormalizable operators (i.e. operators with mass dimension $\lesssim 4$) consistent with the gauge symmetry exist.

1.2 Standard Model - parameter count

The most obvious parameters are the three gauge couplings g_i. Then there is of course the Higgs quartic coupling λ and the Higgs mass parameter m_H (defining the negative quadratic term $-m_H^2|\Phi|^2$). It is not so easy to count the independent Yukawa couplings contained in the three terms

$$\sum_{a,b=1}^{3} \left(\lambda^u_{ab}\, \overline{q}^a_L \Phi^* u^b_R + \lambda^d_{ab}\, \overline{q}^a_L \Phi d^b_R + \lambda^e_{ab}\, \overline{l}^a_L \Phi e^b_R \right) \;+\; \text{h.c.} \tag{1.15}$$

The reader should check that these and only these terms are G_{SM}-invariant, given the general Yukawa-term structure displayed in (1.11). One frequently sees the notation (suppressing generation indices)

$$\overline{q}_L \tilde{\Phi} u_R \qquad \text{with} \qquad \tilde{\Phi}_\alpha = \epsilon_{\alpha\beta}(\Phi^\beta)^* \tag{1.16}$$

for the first term above. This is necessary if one wants to read (1.15) in terms of $SU(2)$ matrix notation. If one simply says that 'group indices are left implicit' (as we do), writing Φ^* is sufficient. Of course, we could also have avoided the explicit appearance of Φ^* in (1.15) altogether by exchanging it with its complex conjugate, implicit in 'h.c.' This is a matter of convention and the form given in (1.15), (1.16) is close what most authors use.

Maybe the easiest way to count the parameters in (1.15) is to think in terms of the low-energy theory with Φ replaced by its VEV. Then the above expression contains three 3×3 complex mass matrices. Furthermore, these mass matrices relate six independent sets of fermions (since the first term only contains u_L and the second only d_L). Thus, the matrices can be diagonalized using bi-unitary transformations - i.e. a basis change of the fermion fields. We are then left with $3 \times 3 = 9$ mass parameters for three sets of up- and down-type quarks and three leptons.

However, the $SU(2)$ gauge interactions give rise to a term in

$$\sum_{a=1}^{3} \overline{q}^a_L \not{D} q^a_L \tag{1.17}$$

which contains both u_L and d_L. It originates in the off-diagonal terms of $\sigma^{1,2}$ which are contained in $\not{D}$. In this u_L/d_L term, the unitary transformation used above does not cancel and a physical

[2] We also note that a slightly different convention, $v \to v'$, with $\Phi_2 = (v' + h)/\sqrt{2}$ and hence $v' = \sqrt{2}v \simeq 246$ GeV is also widely used. Ours has the advantage that $m_t \simeq v$.

3×3 matrix describing 'flavor changing charged currents' (the CKM matrix) is left.[3] Let us write the relevant term symbolically as

$$\sum_{a,b=1}^{3} \overline{u}_L^a \gamma^\mu U^{ab} d_L^b \,. \tag{1.18}$$

The matrix U arises as the product of two unitary matrices from the bi-unitary transformations above. Hence it is unitary.

It will be useful to pause and think more generally about parameterizing a unitary $n\times n$ matrix U. First, $UU^\dagger$ is hermitian, so setting this matrix equal to $\mathbb{1}$ imposes n^2 real constraints. Since U has $2n^2$ real parameters, n^2 parameters are left. Next, recall that orthogonal matrices have $n(n-1)/2$ real parameters or rotation angles. Thus, since unitary matrices are a superset of orthogonal matrices, we may think of characterizing them by $n(n-1)/2$ angles and $n^2 - n(n-1)/2$ phases.[4] Now, in our concrete case, we are free to transform our unitary matrix (in an n-generation Standard Model) according to

$$U \to D_u U \overline{D}_d \,, \tag{1.19}$$

where $D_{u,d}$ are diagonal matrices made purely of phases. This is clear since we may freely rephase the fields u_L^a and d_L^a (together with their mass partners u_R^a and d_R^a – to keep the masses real). The rephasing freedom of (1.19) can be used to remove $2n-1$ phases from U. The '-1' arises since one overall common phase of D_u and D_d cancels and hence does not affect U. So we are left with $n^2 - n(n-1)/2 - (2n-1) = (n-1)(n-2)/2$ physically significant phases.

Now we return to U as part of our Standard Model lagrangian with real, diagonal mass fermion mass matrix. Here $n = 3$ and, according to the above, the CKM matrix has 3 real "mixing angles" and one complex phase (characterizing CP violation in the weak sector of the Standard Model). For more details, see e.g. [8], Chapter 11.3.[5]

This brings our total parameter count to $3+2+9+4 = 18$. However, we are not yet done since we completely omitted a whole general type of term in gauge theories, the so-called topological or θ-term

$$\mathcal{L} \supset \theta\, \mathrm{tr} F \wedge F \sim \theta\, \epsilon^{\mu\nu\rho\sigma} F^a_{\mu\nu} F^a_{\rho\sigma} \,. \tag{1.20}$$

Most naively, this adds 3 new parameters, one for each factor group. However, these terms are total derivatives. Thus, they are invisible in perturbation theory and do not contribute to the Feynman rules. In the non-abelian case, there exist field gauge configurations localized in space and time (called instantons) for which

$$\int \mathrm{tr} F \wedge F \,. \tag{1.21}$$

is non-zero. We will return to them in more detail later. For the $U(1)$, such configurations do not exist, which severely limits the potential observability of the θ term in $U(1)$ gauge theories.

[3]These flavor changing currents correspond to vertices with a (charged) W boson and two left-handed fermions with different flavor (one up and one down, either both from the same or from different generations).

[4]This is not a prove. One needs to show that such a parameterization in terms of angles and phases exists. We will not touch the interesting subject of parameterizations of unitary matrices.

[5]For a broader discussion of C, P, CP and its violation see e.g. [12] and the list of reviews given therein. A further useful set of lecture notes is [13].

Furthermore, and maybe most importantly, the θ term is precisely of the type that the non-invariance of the fermionic path integral measure induces if chiral fermion fields are re-phased. Thus, in the presence of charged fermions without mass terms (or analogous Yukawa-type couplings preventing a re-phasing) such θ parameters are unphysical. The upshot of a more detailed analysis in the Standard Model case (where some but not all conceivable fermionic mass terms are present) is that the $SU(2)$ and $U(1)$ θ terms are unobservable (see e.g. [14]) but the QCD θ term is physical (for some recent discussion non-trivial questions in this context see e.g. [15, 16]). If one goes beyond the Standard Model by adding more fields or even just higher-dimension operators, the electroweak θ terms may become physical.

A non-zero value of θ_{QCD} breaks CP. This is directly visible from the ϵ tensor in its definition as well as from its equivalence (through re-phasing) to complex fermion mass parameters.[6] Now, if CP were broken at the $\mathcal{O}(1)$ level by the theory of strong interactions and if light quark masses where $\sim$GeV, one would expect an $\mathcal{O}(1)$ (in GeV units) electric dipole moment of the neutron to be present. However, even beyond the suppression by the light quark masses $\sim 10^{-3}$ GeV, the dipole moment is experimentally known to be *extremely* small. The detailed analysis of this bound implies roughly $\theta_{QCD} < 10^{-10}$.

In any case, we now arrived at our final result of 19 parameters. However, the status of these parameters is very different. Most notably, 18 of them correspond to dimension-4 (or marginal) operators, while one – the Higgs mass term – is dimension-2 and hence relevant. The latter term refers to ‘relevant in the IR’.

Let us try to make the same point from a more intuitve and physical perspective: Since the theory is renormalizable, one can imagine studying it at a very high energy scale, $E \gg v \sim m_H$. At this scale the Higgs mass is entirely unimportant and we are dealing with a theory of massless fields characterized by 18 dimensionless coupling constants. Classically, this structure is scale invariant since only dimensionless couplings are present. At the quantum level, even without the Higgs mass term, this scale invariance is badly broken by the non-zero beta-functions, most notably of the gauge couplings. Indeed the gauge couplings run quite significantly and, for example, in the absence of the $|\Phi|^2$ term QCD would still confine at about $E \sim 1$ GeV and break the approxinate scale invariance completely.

However, this ‘high-scale’ Standard Model described above is very peculiar in the following sense: One perfectly acceptable operator, $-m_H^2\Phi^2$, is missing entirely. More precisely, if we characterize the theory at a scale μ by dimensionless couplings, e.g. $g_i^2(\mu)$, $\lambda(\mu)$ etc., then we should include a parameter $m_H^2(\mu)/\mu^2$. If we start at some very high scale (e.g. the Planck scale $M_P \sim 10^{18}$ GeV – more on this point later), then this parameter has to be chosen extremely small,

$$m_H^2(\mu)/\mu^2 \sim 10^{-32} \qquad \text{at} \qquad \mu \sim M_P\,, \tag{1.22}$$

to describe our world. Indeed, running down from that scale it keeps growing as $1/\mu^2$ until, at about $\mu \sim 100$ GeV, it starts dominating the theory and completely changes its structure. This is our first encounter with the hierarchy problem, which we will discuss in much more detail below.

[6]Recall that, at the lagrangian level, charge conjugation is related to complex conjugation. In particular, it is broken by complex lagrangian parameters which can not be removed by field redefinitions.

1.3 Effective field theories - cutoff perspective

In this course, we assume familiarity with basic QFT. The language of (low-energy) effective field theory can be viewed as an important part of QFT and hence many readers will be familiar with it. Nevertheless, since this subject is of such an outstanding importance for what follows, we devote some space to recalling the most fundamental ideas of effective field theory (EFT). In addition to chapters in the various QFT books already mentioned, the reader will be able to find many sets of lecture notes devoted specifically to the subject of EFTs, e.g. [17, 18].

To begin, let us assume that our QFT is defined with some UV cutoff Λ_{UV} (and, if one wants, in finite spatial volume $\sim 1/\Lambda_{IR}$), such there can be no doubt that we are dealing with a conventional quantum mechanical system. Of course, the larger the ratio $\Lambda_{UV}/\Lambda_{IR}$, the more degrees of freedom this system has. The possible IR cutoff will not be relevant for us and we will not discuss it further. The best example for UV cutoff (though not very practical in perturbative calculations) is presumably the lattice cutoff. It is e.g. well established that this leads to a good description of gauge theories, including all perturbative as well as non-perturbative effects. Next, it is also well-known and tested in many cases that the lattice regularization can be set up in such a way that Poincare-symmetry is recovered in the IR. Of course, we could use Poincare-invariant cutoffs (e.g. dimensional regularization, Pauli-Villars or even string theory) from the beginning, but the lattice is conceptually simpler and more intuitive. Thus, we will be slightly cavalier concerning this point and assume that we can disregard Poincare-breaking effects in the IR of our system.

As a result (and here we clearly assume a large amount of non-trivial QFT intuition to be developed by reading standard texts) our low-energy physics can be characterized by an action of the symbolic structure (focussing on the gauge theory case)

$$S = \int d^4 x \left(\frac{1}{2g^2} \mathrm{tr} F^2 + \frac{\theta}{8\pi^2}\, \mathrm{tr} F \wedge F + \frac{c_1}{\Lambda^4} \mathrm{tr} F^4 + \frac{c_2}{\Lambda^4} (\mathrm{tr} F^2)^2 + \cdots \right) , \tag{1.23}$$

where $\Lambda \equiv \Lambda_{UV}$ is our cutoff scale. In other words, we expect that generically all terms allowed by the symmetries are present and that, on dimensional grounds, whenever a dimensionful parameter is needed, it is supplied by the the cutoff scale Λ. At low energies, only terms not suppressed by powers of Λ will be important, hence we will always encounter renormalizable theories in the IR. The relevance of terms in IR decreases as their mass dimension grows. This is obvious if one thinks, e.g., in terms of the contribution a given operator makes to a 4-gluon-amplitude: The first term in (1.23) will contribute $\sim g^2$; the third will contribute $\sim g^4 k^4/\Lambda^4$. Clearly, at small typical momentum k, only the first term is important. To see this explicitly one needs to split off the propagator from the first term and to rescale $A_\mu \to g A_\mu$. The lagrangian will then contain terms of the type

$$A\, \partial^2 A \,+\, g\, A^2\, \partial\, A \,+\, g^2\, A^2 \,+\, g^4\, (c_1/\Lambda^4)\, (\partial A)^4 + \cdots \tag{1.24}$$

The numerical coefficients in (1.23) depend on the details of the regularization (e.g. the lattice model) or, if one wants to think of this more physically, of the UV completion at the scale Λ. Specifically, while we assume that this change from the QFT to some finite UV theory occurs at the scale Λ, this transition can depend on many discrete and continuous choices. This will be reflected in the values of g, θ and the c_i. Some of these terms can hence be unusually large or small and this can to a certain extent overthrow the ordering by dimension advertised above. However, in the mathematical limit $k/\Lambda \to 0$, the power of k/Λ is expected to win over numerical prefactors. An exception arises if one coefficient is exactly zero. This possibility will be very important and we will return to this point.

Let us add to our gauge theory example given above the apparently much simpler example of a real scalar field (symmetric under $\phi \to -\phi$):

$$S = \int d^4\,x \left(c_0 \Lambda^2 \phi^2 + \frac{1}{2}(\partial\phi)^2 - \lambda\phi^4 + \frac{c_1}{\Lambda^2}\phi^6 + \cdots \right) . \tag{1.25}$$

The key novelty is that we have a term proportional to a positive power of Λ (a relevant operator). In the gauge theory case, the most important operators were merely marginal. Morover, this term is a mass term and for $c_0 = \mathcal{O}(1)$ the EFT below the scale Λ is simply empty. Thus, we must assume that a very particular UV completion exists which allows for either $c_0 = 0$ (for some qualitative reason) or at least for the possibility to *tune* this coefficient to a very small value, $c_0 \ll 1$. We now see that this has some similarity to the Standard Model, where (assuming that the Standard Model continues to be the right theory above the TeV-scale), a similar tuning might be needed to keep m_H^2 small.

Arguing that there is a 'tuning' or 'fine-tuning problem' based only on the above is not very convincing. One of the reasons is that we were vague about the UV completion at the scale Λ. It appears possible that the right UV completion will effortlessly allow for $c_0 \ll 1$ or maybe naturally predict such a small value. Indeed, we have to admit right away that we will not be able to rule this out during this whole course. But we will try to explain why many researchers have remained pessimistic concerning this option.

1.4 Effective field theories - $\mathbf{QFT_{UV}}$ vs. $\mathbf{QFT_{IR}}$

To do so, we will now modify the use of the word effective field theory: In the above, we assumed some finite (non-QFT) UV completion and called EFT what remains of it in the IR. Now, we want to start with some QFT in the UV (to be itself regularized or UV-completed at even higher scales) and consider how it transits to another QFT (which we will call EFT) in the IR. The simplest way in which this can happen is as follows: Let our QFT_{UV} contain a particle with mass M and focus on the physics at $k \ll M$. In other words, we 'integrate out' the heavy (from the IR perspective) particle and arrive at a theory we might want to call QFT_{IR} – our low-energy EFT.

Let us start with a particularly simple example (borrowed from [17], where a much more detailed discussion of the EFT language can be found):

$$\mathcal{L} = \overline{\psi} i \partial\!\!\!/ \psi - m\overline{\psi}\psi - \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}M^2\phi^2 + y\phi\overline{\psi}\psi - \frac{\lambda}{4!}\phi^4 . \tag{1.26}$$

We assume $m \ll M \ll \Lambda$ and we have already ignored all terms suppressed by Λ. The above lagrangian is renormalizable, such that we are indeed allowed to take the limit $\Lambda \to \infty$ and consider (1.26) (with parameters fixed at some scale $\mu_1 \gg M$) as the definition of our theory. We are interested in the EFT at μ_2 with $m \ll \mu_2 \ll M$.

The correct procedure ('running and matching') would be as follows: One writes down the most general lagrangian $\mathcal{L}_{EFT}$ for ψ at the scale μ_2 and calculates (at some desired loop order) a sufficiently large set of observables (e.g. mass, 4-point-amplitude etc.). Then one calculated the same observables using the full theory defined by (1.26). This includes tree level diagrams and loops involving ϕ as well as the renormalization group (RG) evolution. Finally, one determines the parameters of $\mathcal{L}_{EFT}$ such that the two results agree.

Our course is not primarily about EFTs and we will take a shortcut. First, we set $\lambda = 0$ since it will not be essential in what we have to say. Second, we integrate out ϕ classically: We ignore the $(\partial\phi)^2$ term since we are at low energies and we extremize the relevant part of $\mathcal{L}$ with respect to ϕ:

$$\frac{\delta}{\delta\phi}\left(-\frac{1}{2}M^2\phi^2 + y\phi\overline{\psi}\psi\right) = 0 \qquad \Rightarrow \qquad \phi = \frac{y}{M^2}\,\overline{\psi}\psi\,. \tag{1.27}$$

Inserting this back into our lagrangian we obtain

$$\mathcal{L}_{EFT} = \overline{\psi}i\partial\!\!\!/\psi - m\overline{\psi}\psi + \frac{y^2}{2M^2}(\overline{\psi}\psi)^2 + \cdots\,. \tag{1.28}$$

Finally, we calculate loop corrections involving the heavy field ϕ to all operators that potentially appear in $\mathcal{L}_{EFT}$. In this last step, the correction which is most critical for us is the mass (or more generally the self energy correction) for ψ, cf. Fig. 1.

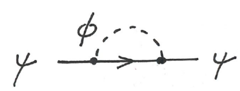

Figure 1: One-loop fermion self energy in the Yukawa theory.

Dropping all numerical prefactors, this gives (for details see e.g. [6])

$$\Sigma(p\!\!\!/) \sim y^2 \int d^4\,k \frac{-k\!\!\!/ + m}{(k^2+m^2)[(k+p)^2+M^2]}\,. \tag{1.29}$$

After summing, in the standard way, all such self-energy corrections to the propagator, one obtains

$$\frac{i}{p\!\!\!/ - m - \Sigma(p\!\!\!/)}\,. \tag{1.30}$$

This resummed propagator can be viewed as a function of the matrix-valued argument $p\!\!\!/$. Its pole then determined the corrected mass $m_c = m + \delta m$. This can be made explicit by Taylor expanding $\Sigma(p\!\!\!/)$ around $p\!\!\!/ = m_c$:

$$\Sigma(p\!\!\!/) = \Sigma(m_c) + \Sigma'(m_c)(p\!\!\!/ - m_c) + \frac{1}{2}\Sigma''(m)(p\!\!\!/ - m_c)^2 + \cdots\,. \tag{1.31}$$

Now the propagator takes the form

$$\frac{i}{(p\!\!\!/ - m_c) - \Sigma'(m_c)(p\!\!\!/ - m_c) + \cdots} \qquad \text{with} \qquad m_c = m + \Sigma(m_c)\,. \tag{1.32}$$

As usual in perturbation theory we estimate

$$\delta m = \Sigma(m)\,. \tag{1.33}$$

Thus, we need to evaluate (1.29) for $\not{p} = m$. Introducing a cutoff Λ, we have in total three scales: m, M and Λ. We see right away that the (naively possible) linear divergence arising from the term $\sim \not{k}$ will vanish on symmetry grounds as long as our cutoff respects Lorentz symmetry. Morover, the contribution from the term $\sim \not{k}$ in the regime $k \sim m \ll M$ is suppressed by $1/M^2$. Thus, we may disregard the term $\sim \not{k}$ altogether.

We may then focus on the term $\sim m$. It gets a small contribution from the momentum region $k \lesssim M$ and is log-divergent for $k \gg M$. We can finally conclude that the leading result for the mass correction extracted from (1.29) must be proportional to m. Any enhancement beyond this can at best be logarithmic, but still proportional to m. For the moment this is all we need: We learn that (1.28) is the right lagrangian after the replacement

$$m \;\to\; m_{EFT} \equiv m_c \equiv m + \delta m \equiv m(1 + y^2 \times \mathcal{O}(1))\,. \tag{1.34}$$

Here '$\mathcal{O}(1)$' may include a logarithmic cutoff dependence, like in particular $\ln(\Lambda/M)$. Moreover, as noted earlier, we may define our theory at a finite scale $\mu_1 \gg M$. Then the log-divergence is traded for $\ln(\mu_1/M)$.

We could have argued the same without even drawing any Feynman diagram: Indeed, writing our model in terms of left- and right-handed spinors,

$$\overline{\psi}\psi = \overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L\,, \tag{1.35}$$

one sees immediately that for $m = 0$ it possesses the $\mathbb{Z}_2$ symmetry

$$\psi_L \to \psi_L\,, \qquad \psi_R \to -\psi_R\,, \qquad \phi \to -\phi\,. \tag{1.36}$$

The mass term $\sim m$ breaks this symmetry. Thus, we expect that both the UV theory and the EFT regain this symmetry in the limit $m \to 0$. The loop correction δm of

$$m_{EFT} = m + \delta m \tag{1.37}$$

must hence itself be proprtional to m. The punchline is that, integrating out ϕ, does not clash with the lightness (or masslessness) of ψ.

It is interesting and important to develop this language further by considering the low-energy EFT of the Standard Model below the scale of Higgs, W and Z-bosons or the pion EFT below the confinement scale Λ_{QCD}. We leave it to the reader to explore this using the vast literature.

1.5 The Standard Model as an effective field theory

Let us now apply the above language to the Standard Model. We first assume that a finite cutoff $\Lambda \gg$ TeV is present and that the Standard Model is the effective theory valid below this cutoff. At the moment, we allow this cutoff to either be the scale at which the framework of QFT becomes insufficient (string or some other fundamental cutoff scale) or, alternatively, the scale at which the Standard Model is replaced by a different, more fundamental, ultraviolet QFT. It is natural to view Λ as our main dimensionful parameter and organize the langrangian as

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_2 + \mathcal{L}_4 + \mathcal{L}_5 + \mathcal{L}_6 + \cdots &(1.38)\\ &= c_0\Lambda^2|\Phi|^2 - |D\Phi|^2 - \lambda|\Phi|^4 + \mathcal{L}_4' + \mathcal{L}_5 + \mathcal{L}_6 + \cdots &(1.39)\end{aligned}$$

Here $\mathcal{L}_4'$ is our familiar renormalizable Standard Model lagrangian without the Higgs terms, which we displayed explicitly. We also have $m_H^2 = c_0\Lambda^2$ and we note that $c_0 \ll 1$ is necessary (with the smallness depending on how high Λ actually is). But our discussion in the previous section has not lead to an unambiguous conclusion about whether this should be viewed as a problem.

Since we now think of the Standard Model as of an EFT, we included terms of mass dimension 5, mass dimension 6, and so on. It turns out that, at mass dimension 5, the allowed operator is essentially unique (up to the flavor stucture). We write it down for the case of a single family and using a two-component (Weyl) spinor doublet l_α.

$$l_L = \begin{pmatrix} l_\alpha \\ 0 \end{pmatrix}. \tag{1.40}$$

This so-called Weinberg-operator then takes the form

$$\mathcal{L}_5 = \frac{c}{\Lambda}(l\cdot\Phi)^2 + \text{h.c.} = \frac{c}{\Lambda} l_i^\alpha l_{\alpha j}\epsilon^{ik}\epsilon^{jl}\Phi_k\Phi_l + \text{h.c.} \tag{1.41}$$

Here we used the fact that two Weyl spinors can form a Lorentz invariant as

$$\psi^\alpha\psi_\alpha = \epsilon^{\alpha\beta}\psi_\beta\psi_\alpha\,, \tag{1.42}$$

where the ϵ tensor appears in its role as an invariant tensor of $SL(2,\mathbb{C})$. The ϵ-tensors in (1.41) appear in their role as invariant tensors of the $SU(2)$ factor G_{SM} and allow us to combine two doublets (Higgs and leptons) into a singlet.

Now, since

$$\langle\Phi\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.43}$$

the low energy effect of the above operator is to give mass to the upper component of the lepton doublet, i.e. to the neutrino:

$$\mathcal{L}_5 = \frac{cv^2}{\Lambda}\nu^\alpha\nu_\alpha + \text{h.c.} \tag{1.44}$$

Writing the neutrino as a Majorana rather than a Weyl fermion, this becomes the familar Majorana mass term. Introducing three families, the constant c is promoted to a 3×3 matrix c_{ab}.

Given our knowledge that neutrino masses are non-zero and (without going into the non-trivial details of the experimental situation) are of the order $m_\nu \sim 0.1$ eV, an effective field theorist can interpet the situation as follows: The neutrino mass measurements represent the detection of the first higher-dimension operator of the Standard Model as an EFT. As such, they determine the scale Λ via the relation (assuming $c = \mathcal{O}(1)$)

$$m_\nu \sim v^2/\Lambda \qquad \Rightarrow \qquad \Lambda \sim 3 \times 10^{14} \text{ GeV}\,. \tag{1.45}$$

On the one hand, this is discouragingly high. On the other hand, it is significantly below the (reduced) Planck scale of $M_P \simeq 2.4 \times 10^{18}$ GeV. It is also relatively close to, though still significantly below, the supersymmetric Grand Unification scale $M_{GUT} \sim 10^{16}$ GeV to which will return later. Let us note that, without supersymmetry, the GUT scale is less precisely defined and one may argue that the UV scale derived from the Weinberg operator above is actually intriguingly close to such a more general GUT scale.

It is very remarkable that the Standard Model with the Weinberg operator allows for a simple UV completion at the scale Λ. This so-called see-saw mechanism [19] involves (we discuss the one-generation case for simpicity) the addition of just a single massive fermion, uncharged under G_{SM}. The relevant part of the high-scale lagrangian is (in Weyl notation for spinors)

$$\mathcal{L} \supset \beta l \Phi \nu_R - \frac{1}{2} M \nu_R \nu_R + \text{h.c.} \tag{1.46}$$

Integrating out the extra fermion (often referred to as the right-handed neutrino ν_R), one obtains precisely the previously given Weinberg operator with

$$c \sim \beta^2 \qquad \text{and} \qquad \Lambda \sim M\,. \tag{1.47}$$

In other words, the observed neutrino masses behave as

$$m_\nu \sim \beta^2 v^2/M\,. \tag{1.48}$$

As a result, we can make M (and thus Λ) smaller, bringing it closer to experimental tests, at the expense of also lowering β. Of course, one has to be lucky to actually discover ν_R at colliders, given that then β would have to take the rather extreme value of $\sqrt{100 \text{ GeV}/10^{14} \text{ GeV}} \sim 10^{-6}$.

An even more extreme option, which however has its own structural appeal, is to set M to zero. This can be justified, e.g., by declaring lepton number to be a good, global symmetry of the Standard Model (extended by r.h. neutrinos). By this we mean the $U(1)$ symmetry $l \to e^{i\chi} l$, $\nu_R \to e^{-i\chi}\nu_R$. Now the Standard Model has an extra field, the fermionic singlet ν_R (more precisely three copies of it). The first term in (1.46) is just another Yukawa coupling (given here in Weyl notation, but otherwise completely analogous to the e.g. the electron Yukawa term). The second term is missing. This version of the Standard Model, extended by r.h. neutrinos, is again a renormalizable theory and it can account for the observed neutrino masses. The latter do not arise from the see-saw mechanism sketched above, but correspond simply to a tiny new Yukawa coupling. In this case $\beta \sim m_\nu/v \sim 10^{-12}$, which may be perceived as uncomfortably small. The second smallest coupling would be that of the electron, $\beta_e \sim 0.5 \text{ MeV}/v \sim 10^{-5}$.

At mass dimension 6, there are many further terms that can be added to $\mathcal{L}_{SM}$. For example, any term of $\mathcal{L}_4$ can simply be multiplied by $|\Phi|^2$. The arguably most interesting terms are the 4-fermion-operators. They include terms like (again in Dirac notation)

$$\mathcal{L}_6 \supset \sum_{ijkl} c_{ijkl} (\overline{\psi}_i \psi_j)(\overline{\psi}_k \psi_l) \tag{1.49}$$

as well as similar operators involving gamma matrices. Even with the restriction by gauge invariance, there are many such terms and we will not discuss them in any detail. Crucially, many of them are very strongly constrained experimentally. First, if one does not impose the global symmetries of lepton and baryon number (the latter being the obvious generalization of the first, which was introduced above), some of these operators induce proton decay. This would push Λ up beyond 10^{16} GeV. But even imposing baryon and lepton number as additional selection rules[7] for (1.49), strong constraints remain. These are mostly due to so called flavor-changing neutral currents (the analogues of the flavor-changing charged currents mentioned earlier) and to lepton flavor violation (e.g. the decay $\mu^- \to e^+ + 2e^-$). Such constraints push Λ to roughy 10^3 TeV. Of course, the new physics scale can be much lower if the relevant new physics has the right 'flavor properties' not to clash with data.

1.6 The electroweak hierarchy problem

Now we come in more detail to what is widely considered the main problem of the Standard Model as an effective theory: the smallness of the Higgs mass term. So far, we have only pointed out that, in the EFT approach with cutoff Λ, it is natural to write

$$m_H^2 \sim c_0 \Lambda^2 \,. \tag{1.50}$$

We have many reasons to think that Λ is large compared to the weak scale, implying $c_0 \ll 1$. The main question hence appears to be whether we can invent a more fundamental theory at scale Λ in which $c_0 \ll 1$ can be understood.

Let us first give a very simple argument (though possibly not very strong) why this is not easy. Namely, consider the theory as given by a classical lagrangian at Λ and ask for low-energy observables. The most obvious is maybe a gauge coupling,

$$\alpha_i^{-1}(\mu) \simeq \alpha_i^{-1}(\Lambda) + \frac{b_i}{2\pi} \ln\left(\frac{\Lambda}{\mu}\right) + \mathcal{O}(1) \,, \tag{1.51}$$

where we restricted attention to the one-loop level. The relevant diagrams are just the self-energy diagrams of the corresponding gauge boson with scalars, fermions and (in the non-abelian case) gauge bosons running in the loop. We see that, for $\Lambda \gg \mu$, the correction becomes large, but it grows only logarithmically (corresponding to the logarithmic divergence of the relevant diagrams and the vanishing mass dimension of the coupling). By contrast, for the Higgs mass we find

$$m_H^2(\mu) = m_H^2(\Lambda) + \frac{c_H}{16\pi^2}\Lambda^2 + \mathcal{O}(\Lambda^0) \,, \tag{1.52}$$

with c_H a coupling-dependent dimensionless parameter to be extracted from diagrams like those in Fig. 2. We see that, suppressing $\mathcal{O}(1)$ coefficients and disregarding the logarithmic running of the dimensionless couplings between μ and Λ, we have $c_H = \lambda + \lambda_t^2 + g_2^2 + \cdots$.

[7] To be precise, the two corresponding $U(1)$ symmetries, known as $U(1)_B$ and $U(1)_L$ are so-called accidental symmetries of the Standard Model. This means that, given just gauge symmetry and particle content, and writing down allowed renormalizable operators, these symmetries are automatically preserved at the classical level. It is hence not unreasonable to assume that they hold also in certain UV completions and may constrain 4-fermion operators.

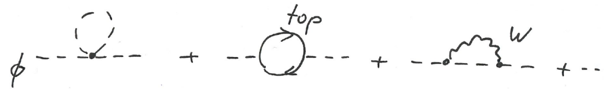

Figure 2: Contributions to the Higgs self energy.

Thus, c_H is an $\mathcal{O}(1)$ number and if we set $\Lambda = 1$ TeV, only an $\mathcal{O}(1)$ cancellation between the two terms on the r.h. side of (1.52) is required to get the right Higgs mass parameter of the order of $(100 \text{ GeV})^2$. Things are actually a bit worse since there is a color factor of 3 coming with the top and other numerical factors. But, much more importantly, we can *not* simply declare 1 TeV to be the scale where our weakly coupled QFT breaks down and some totally unknown new physics (discrete space time, string theory etc.) sets in. One but not the only reason is the issue of flavor-changing neutral currents mentioned above. If we take the (still rather optimistic) value $\Lambda \sim 10$ TeV, we already require a compensation at the level of 1% or less between the two leading terms on the r.h. side of (1.52). This starts to deserve the name fine-tuning.

A cautionary remark concerning expressions like $m_H^2(\mu)$ or $m_H^2(\Lambda)$ is in order. Such dimensionful parameters sometimes (not always) have power-divergent loop corrections. The momentum integral implicit in the loop correction is then dominated in the UV and changes by an $\mathcal{O}(1)$ factor if the regularization procedure changes. This is in contrast to e.g. $\alpha^{-1}(\mu)$ which is, at leading order, independent of how precisely the scale μ is defined. One can see that most easily by noting that $\ln(\Lambda/\mu)$ does not change significantly in the regime $\Lambda/\mu \gg 1$ if Λ or μ are multiplied by, say, a factor of 2. Thus, a possibly less misleading way to write (1.52) is

$$m_H^2 = m_{H,0}^2 + \frac{c_H}{16\pi^2}\Lambda^2 + \mathcal{O}(\Lambda^0)\,, \tag{1.53}$$

Here m_H^2 is, by definition, the value of this operator in the IR and $m_{H,0}^2$ is the bare or classical value in the UV lagrangian.

Still, the fine-tuning argument is not very convincing since, in (1.53), the two crucial terms between which a cancellation is required both depend on the cutoff or regularization used. For example, in dimensional regularization with minimal subtraction, the second term is simply zero and no cancellation appears necessary. Now this is clearly unphysical, but one may entertain the hope that some physical cutoff with similar features will eventually be established, defining a UV theory with a 'naturally' small m_H^2 in spite of large Λ.

But a much more technical and stronger argument making the fine-tuning explicit can be given. We make it using a toy model, but the relevance to the Standard Model will be apparent. The toy model is in essence something like 'the inverse' of the Yukawa model of (1.26). There, we considered the mass correction a fermion obtains when a heavy scalar is integrated out. We found that no large correction to the small fermionic mass arises. Now consider (again following [17]),

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 + \overline{\psi}(i\partial\!\!\!/ - M)\psi + y\phi\overline{\psi}\psi\,. \tag{1.54}$$

We literally simply renamed $m \leftrightarrow M$, having of course in mind that now $m \ll M$. As before, we will not go through a careful procedure of 'running and matching' to derive the low-energy EFT, but take the shortcut of integrating out the heavy field classically and adding loop corrections to the low-energy lagrangian terms.

Since the fermion appears only quadratically in the action, its equations of motion are solved by $\psi = 0$ for any field configuration $\phi(x)$. Hence, the first step consists in just dropping all terms with ψ. When considering loops, we focus only on corrections to the scalar mass proportional to y^2, finding

$$\mathcal{L}_{EFT} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m_{EFT}^2\phi^2 - \frac{\lambda}{4!}\phi^4 + \cdots \,, \tag{1.55}$$

with

$$m_{EFT}^2 \simeq m^2 + \frac{cy^2}{16\pi^2}\int_0^{\Lambda^2} k^2\, d(k^2)\, \frac{\mathrm{tr}(\not{k} - M)^2}{(k^2 + M^2)^2}\,. \tag{1.56}$$

Here c is a numerical constant and the integral corresponds to the second diagrams of Fig. 2. It is immediately clear that both terms proportional to Λ^2 as well as to M^2 will arise:

$$m_{EFT}^2 \simeq m^2 + \frac{y^2}{16\pi^2}\left(c_1\Lambda^2 + c_2 M^2\ln(\Lambda^2/M^2) + c_3 M^2 + \cdots\right)\,. \tag{1.57}$$

See e.g. [6] for a corresponding analysis in dimensional regularization. (Note that, while a quadratic divergence in 4 dimensions does not show up as a pole at $d = 4$, it corresponds to a logarithmic divergence in 2 dimensions and hence shows up as a pole at $d = 2$.)

Crucially, we now see that if, by some 'UV miracle', the m^2 and Λ^2 terms always cancel to make m_{EFT}^2 very small, the tuning issue still remains: Even a very tiny relative change of M^2 (assuming that $M^2 \gg m_{EFT}^2$), would upset this cancellation. Of course, we can not rule out a UV model where *everything*, including masses of particles at intermediate scale (like our M with $m_{EFT} \ll M \ll \Lambda$) are automatically correctly adjusted to ensure the necessary cancellation (1.57). But now it becomes more apparent how tricky any mechanism accoplishing that would have to be.

Concretely in the Standard Model with a see-saw mechanism for neutrino masses, the scale M might be that of the heavy r.h. neutrino and one has, given the above, a strong argument for fine-tuning. Alternatively, one can of course avoid any such heavy particles (also giving up on Grand Unifcation - see below) and imagine that the Standard Model directly runs into a new theory in the UV where, at some scale Λ, a massless scalar is explained without tuning. I am not aware of any such scenario, but we will nevertheless return to a more detailed discussion of this and related logical possibilities later on.

For now, let us accept that, from an EFT perspective, the Standard Model with UV scale Λ is fine tuned and try to quantify the problem.

1.7 Fine tuning

Let us first emphasize that, having a small (dimensionful or dimensionless) parameter in an EFT is not in itself problematic or related to tuning. Indeed, the electron mass is small, but it comes from a dimensionless Yukawa coupling which only runs logarithmically. Thus, once small in the UV, it will stay small in the IR 'naturally'.

Moreover, the relevant coupling of type

$$\lambda_e \bar{l}_L \Phi e_R \tag{1.58}$$

is forbidden by chiral symmetry transformations, e.g. $e_R \to e^{i\alpha} e_R$. One can view λ_e as a small effect in the UV lagrangian breaking this symmetry. Hence, the above operator will only receive loop corrections proportional to this symmetry breaking effect, i.e. to λ_e itself.

Moreover, the same argument can be made for fermion masses even when they are viewed as dimensionful parameters. We have seen one example in Sect. 1.3. Another example is the Standard Model below the electroweak symmetry breaking scale, where the electron mass term

$$m_e \bar{e} e = m_e \bar{e}_L e_R + \text{h.c.} \tag{1.59}$$

can be forbidden by the global $U(1)$ symmetry $e_R \to e^{i\alpha} e_R$, as above. Hence, there will be no loop corrections driving m_e up to the electroweak scale, given that the tree-level value is small.

Small parameters with this feature are called 'technically natural', a notion due to 't Hooft [22]. More precisely, a small parameter is **technically natural** if, by setting it to zero, the symmetry of the system is enhanced. The crucial point for us is that finding such a symmetry for the Higgs mass term turns out to be difficult if not impossible: One obvious candidate is a shift symmetry, $\Phi \to \Phi + \alpha$, with α =const. But this forbids all non-derivative couplings and hence clashes with the main role the Higgs plays in the Standard Model, most notably with the top-Yukawa coupling, which is $\mathcal{O}(1)$. Nevertheless, attempts to at least alleviate the hierarchy problem using this idea have been made and we will discuss them. Another option is scale-invariance but, once again, the Standard Model as a quantum theory is not scale invariant - couplings run very significantly. Moreover, in the UV, most ideas for how the unification with gravity will work break scale invariance completely. Again, attempts along these lines nevertheless exist and will be discussed. However, at our present 'leading order' level of discussion it is fair to say that the smallness of m_H^2 is probably not technically natural.

Somewhat more vaguely, one may say that the Higgs mass term is unnaturally small. To make this statement more precise, the notion of tuning or **fine tuning** has been introduced. Roughly speaking, a theory is tuned if the parameters in the UV theory (at the scale Λ) have to be adjusted very finely to realize the observed low-energy EFT.

It is not immediately obvious how to implement this in terms of formulae since, as just explained, e.g. the electron mass is known with high accuracy and even a tiny change of the UV-scale Yukawa coupling will lead to drastic disagreement with experiment. The main point one wants to make is that, as we have seen, the smallness of the Higgs mass apparently arises from the **compensation** between two terms,

$$m_H^2 = m_{H,0}^2 + \frac{c_H}{16\pi^2}\Lambda^2 + \cdots . \tag{1.60}$$

Clearly, in such a situation, a small *relative* change, e.g., $m_{H,0}^2$ induces a much larger *relative* change of m_H^2.

A widely used formula implementing this is known as the **Barbieri-Giudice measure** for fine-tuning [23] (see also [24] and, for a modern exposition, [25]):

$$FT = \left| \frac{x}{\mathcal{O}} \frac{\partial \mathcal{O}}{\partial x} \right| = \left| \frac{\partial \ln(\mathcal{O})}{\partial \ln(x)} \right| . \tag{1.61}$$

Here x is the theory parameter and $\mathcal{O}$ the relevant observable. In our case, $x = m_{H,0}^2$ and $\mathcal{O} = m_H^2$ is given by (1.60), such that

$$FT = m_{H,0}^2 \frac{\partial \ln(m_{H,0}^2 + c_H\Lambda^2/(16\pi^2))}{\partial m_{H,0}^2} = \frac{m_{H,0}^2}{m_{H,0}^2 + c_H\Lambda^2/(16\pi^2))} \sim \frac{\Lambda^2/(16\pi^2)}{m_H^2} . \tag{1.62}$$

Here, in the last step, we assumed that $m_H^2 \ll c_H\Lambda^2/(16\pi^2)$, such that $m_{H,0}^2 \sim c_H\Lambda^2/(16\pi^2)$. Moreover, we have used that $c_H = \mathcal{O}(1)$. As already noted earlier, this just formalizes what we said at the intuitve level earlier: The fine tuning is roughly $\Lambda^2/(1\text{ TeV})^2$.

For completeness, we record the natural multi-particle generalization of the Barbieri-Giudice measure. In this more general context, one may call it a 'fine tuning functional', defined as a functional on the space of theories T (following [25]):

$$FT[T] = \sum_{ij} \left| \frac{x_i}{\mathcal{O}_j} \frac{\partial \mathcal{O}_j}{\partial x_i} \right| . \tag{1.63}$$

We also note that our discussion was somewhat oversimplified and less concrete than in [23]. There, the observable was m_Z^2 (this is clearly tied to m_H^2, which is however not directly observable). Furthermore, the UV theory was not some very vague cutoff-QFT but it was a concrete UV QFT (the supersymmetric, in fact even supergravity-extended version of the Standard Model). We will get at least a glimpse of this below.

Unfortunately, the above definition of fine-tuning has many problems. First, it is clearly not reparameteization independent. In other words, it crucially depends on our ad hoc choice of x_i as operator coefficients in a perturbative QFT and of the $\mathcal{O}_i$ as, roughly speaking, particle masses. Thus, one is justified in looking for other, possibly related, definitions. One such alternative definition is probabilistic: Choose a (probability) **measure on the set of UV theories** and ask how likely it is to find a particular low-energy observable to lie in a certain range. For example, we might consider $m_{H,0}^2$ to have a flat distribution between zero and $2\Lambda^2/(16\pi^2)$ (where we also set $c_H = -1$). Then we obtain a small Higgs mass only if $m_{H,0}^2$ happens, by chance, to fall very close to the center point of its allowed range.

To make this quantitative in the sense just outlined, we will use the variable m_H^2 for the Higgs mass squared in any of our statistical set of theories. By contrast, we will denote the concrete Standard Model value by $m_{H,obs.}^2$. Then, we may ask for the probability to find m_H^2 in the interval $[-m_{H,obs.}^2, m_{H,obs.}^2]$, or equivalently $|m_H^2| \lesssim m_{H,obs.}^2$. We obtain

$$p(m_{H,obs.}^2) \simeq \frac{m_{H,obs.}^2}{\Lambda^2/(16\pi^2)} . \tag{1.64}$$

This is just the inverse of the Barbieri-Giudice fine tuning value, confirming at least at some intuitive level that the above definitions make sense. However, it becomes even more apparent that some ad hoc assumptions have come in. In particular, we required a measure on the space of UV theories or UV parameters.

Finally, another ambiguity of the probabilistic view on fine-tuning is related to the choice of the allowed interval of the EFT observable. In the above, things were rather clear since our task was to quantify the problematic *smallness* of the Higgs mass relative to the cutoff. It was then natural to define all theories with $|m_H^2| \lesssim m_{H,\,obs.}^2$ as 'successful'. However, the Higgs mass is by now known rather precisely, $m_h = 125.18 \pm 0.16$ GeV [26], which translates in similarly precise value for m_H^2. If we had defined successful theories as those with $m_{H,\,obs.}$ falling into that interval, a much larger fine-tuning would result. Even much worse, one could consider the very precisely known electron mass in the same way and would find a huge fine tuning of the UV-scale coupling λ_e, in spite of the logarithmic running and the technical naturalness.

Thus, one has to be careful with both definitions and it may well be that the final word about this has not yet been spoken. A suggestion for sharpening the probabilistic perspective is as follows: Consider the manifold of UV couplings (with some measure) and the map to the manifold of observables. On the latter, let $\mathcal{O}_0$ be some **qualitatively** distinguished point, in our case the point of vanishing Higgs mass term. This point is distinguished since it specifies the boundary between two qualitatively different regimes – that with spontaneously broken and unbroken $SU(2)$ gauge symmetry. Let us assume that for any other point $\mathcal{O}$, we can in some way measure the distance to this special point, $|\mathcal{O} - \mathcal{O}_0|$. Now one may say that an *observed* EFT, corresponding to a point $\mathcal{O}_{obs.}$ on the manifold of observables, is fine tuned to the extent that the probability for all theories with

$$|\mathcal{O} - \mathcal{O}_0| < |\mathcal{O}_{obs.} - \mathcal{O}_0| \tag{1.65}$$

is small. In other words, we measure how unlikely it is that a randomly chosen theory falls more closely to the special point $\mathcal{O}_0$ than our EFT under discussion.

1.8 Gravity and the cosmological constant problem

If we include gravity,

$$\mathcal{L}_{SM}[\psi, \eta_{\mu\nu}] \qquad \to \qquad \mathcal{L}_{SM}[\psi, g_{\mu\nu}] + \frac{1}{2} M_P^2 \sqrt{g}\, \mathcal{R}[g_{\mu\nu}] - \sqrt{g}\lambda\,, \tag{1.66}$$

two essential modifications of the discussion above arise. First, we learn that the Higgs mass problem is just one of two instances of very similar hierarchy problems - the other being the cosmological constant problem. Second, gravity sets an upper bound on the cutoff Λ, in a way that sharpens the Higgs mass hierarchy problem.

In more detail, let us start by recalling what we need to know about gravity [27–31]. On the one hand, gravity changes the picture very deeply: The arena of our Standard-Model QFT changes from $\mathbb{R}^4$ (with flat Lorentzian metric) to a Lorentzian manifold with dynamical metric, horizons, singularities in the cosmic past or future, or possibly even with topology change. The causal structure, which is so crucial for the definition of a QFT, becomes dynamical together with the metric. In particular, if one takes the metric itself to be a dynamical quantum field, the quantization of this field depends on the causal structure, which follows from the (then a priori unknown) dynamics of this field itself. Diffeomorphism invariance makes it very hard to define what a local observable in the usual QFT sense is supposed to be. Finally, to just mention one more issue, QFTs are most easily defined in Euclidean metric. But this is extremely problematic in gravity since even for a 4d euclidean manifold $\mathcal{R}[g_{\mu\nu}]$ can take either sign. Thus, fluctuations around a flat euclidean background do not necessarily suppress the weight factor $\exp(-S_E)$ in the path integral, the well-definedness of which hence becomes problematic.

But, on the other hand, one may also ignore most of the deep conceptual problems above and pretend that one has added to the Standard Model QFT just another gauge theory (see e.g. [32]). We can not develop this approach here in any detail but only sketch the results: One expands the metric around flat space,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \,. \tag{1.67}$$

and tries to think of $h_{\mu\nu}$ as a gauge potential, analogous to A_μ. One recalls that

$$D_\mu v_\nu = \partial_\mu v_\nu - \Gamma_{\mu\nu}{}^\rho v_\rho \qquad \text{with} \qquad \Gamma_{\mu\nu}{}^\rho = \frac{1}{2} g^{\rho\sigma}(\partial_\mu g_{\nu\sigma} + \cdots - \cdots) \tag{1.68}$$

and (symbolically, suppressing the index structure)

$$R \sim [D, D] \sim [\partial - \Gamma, \partial - \Gamma] \,. \tag{1.69}$$

From this, it is clear that the gravitation lagrangian takes the form

$$M_P^2 \, [\, h\partial^2 h + h(\partial h)^2 + h^2(\partial h)^2 + \cdots] \,. \tag{1.70}$$

Defining $\kappa \equiv 1/M_P$ and rescaling $h \to \kappa h$, this becomes

$$h\partial^2 h + \kappa h(\partial h)^2 + \kappa^2 h^2(\partial h)^2 + \cdots \,. \tag{1.71}$$

This is already quite analogous to the gauge theory structure (we are thinking of the non-abelian case, but suppress the group and gauge indices for brevity)

$$A\partial^2 A + gA^2\partial A + g^2 A^4 \,. \tag{1.72}$$

The crucial differences are that g is dimensionless and the series of higher terms terminates at the quartic vertex. By contrast, in gravity the coupling has mass dimension -1 and the series goes on to all orders (both from R as well as from the R^2, R^3 terms etc. which have to be added to the lagrangian to absorb all divergences arising at loop level). We will not discuss the technicalities of this – suffice it to say that the Fadeev-Popov procedure and the introduction of ghosts work, at least in principle, as in gauge theories.

We also recall that, for any observable that we can calculate in perturbation theory, the expansion reads

$$c_0 + c_1\kappa\Lambda + c_2\kappa^2\Lambda^2 + \cdots \tag{1.73}$$

on dimensional grounds. From this we see that we have to expect power divergences and that higher loops are more and more divergent, consistent with the well-known fact that quantum gravity is perturbatively non-renormalizable.

Finally, coming closer to our main point, we remember that $g_{\mu\nu}$ or, in our approach $h_{\mu\nu}$ appears also in $\mathcal{L}_{SM}[\psi, g_{\mu\nu}]$. Since, as we know, the energy momentum tensor is defined as the variation of S_{SM} with respect to $g_{\mu\nu}$ at the point $g_{\mu\nu} = \eta_{\mu\nu}$, it is clear that the leading order coupling of h with matter is given by

$$\mathcal{L} \supset \kappa h_{\mu\nu} T^{\mu\nu} \,. \tag{1.74}$$

This is, once again, completely analogous to the gauge theory coupling to matter via

$$\mathcal{L} \supset g A_\mu^A j_A^\mu \qquad \text{with, e.g.} \qquad \overline{\psi}\gamma^\mu T_A \psi \,. \tag{1.75}$$

What is essential for us is that the cosmological constant term gives rise to an energy momentum tensor

$$T^{\mu\nu} \sim \eta^{\mu\nu}\lambda \,. \tag{1.76}$$

If λ is non-zero, then this gives rise to a non-zero source (a tadpole term) for the metric (gauge) field $h_{\mu\nu}$:

$$\mathcal{L} \supset \kappa h_{\mu\nu}\eta^{\mu\nu}\lambda \,. \tag{1.77}$$

The meaning of the word tadpole in this context becomes obvious if one considers the above as a tree-level diagrammatic effect and adds the first loop correction (due e.g. to a scalar particle minimally coupled to gravity). This is illustrated in Fig. 3.

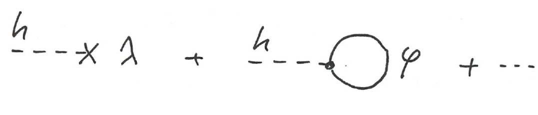

Figure 3: Tree level and loop effect of the cosmological constant term on the metric field $h_{\mu\nu}$.

One may think of the loop diagram in Fig. 3 as a correction to λ, in direct analogy to the loop corrections to the Higgs mass from integrating out heavy particles which we discussed before. Thus, in analogy to e.g. (1.53) and renaming our original cosmological term in the tree-level action to λ_0, we have

$$\lambda = \lambda_0 + \frac{c_\lambda}{16\pi^2}\Lambda^4 \,. \tag{1.78}$$

The coefficient c_λ does not include a small coupling constant but is merely proportional to the sum of bosonic and fermionic degrees of freedom (one may interpet this sign difference either as being due to the usual 'minus' for each fermion loop or as the negative sign of the vacuum energy of the fermionic harmonic oscillator. Related to this, one can of course interpret the divergence as a sum over the vacuum energies of the oscillators corresponding to free field momentum modes.

Famously, if one compares the observed value of the vacuum energy,

$$\lambda \simeq (2.2\,\text{meV})^4 \,, \tag{1.79}$$

with the expectation from (1.78) based on $\Lambda = M_P \simeq 2.4 \times 10^{18}$ GeV, one finds a mismatch (i.e. a required fine tuning) of 10^{120}. As in the Higgs mass case, there are caviats to this argument: Indeed, the value of the loop correction depends completely on the UV regularization and one may imagine schemes where it would simply be zero. Also, as in Higgs case, there are couterarguments to this suggestion. Indeed, any massive particle contributes to the loop correction in a way that depends on its mass. Thus, the observed value changes dramatically if, e.g., the mass of the heavy r.h. neutrino needed in the see-saw mechanism changes.

To see this more explicitly, it is useful to give an explicit covariant formula for the one-loop correction to λ (see problems for a derivation). For a single real scalar and in euclidean signature, one has

$$\delta\lambda = \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\,\ln(k^2 + m^2) = c_0\Lambda^4 + c_1\Lambda^2 m^2 + \cdots \,. \tag{1.80}$$

We see that even the subleading term proportional to the mass is still also proportional to Λ^2 and hence huge. In fact, this is true even for the light particles of the Standard Model. Furthermore, there are effects due to the Higgs potential, the non-perturbative gluon-condensate of QCD and from all couplings (which enter at the two and higher-loop level). Thus, the case for an actual fine-tuning appears to be very strong indeed. Clearly, the amount of fine tuning may be strongly reduced compared to what we just estimated: We could add to the Standard Model heavy bosons and fermions, such that above a certain mass scale M the number of fermions and bosons is equal and at least the leading Λ^4 term disappears.

This last idea turns out to work much better than expected. It is realized in a systematic way in supersymmetry (SUSY) or supergravity (SUGRA). It still does not solve the cosmological constant problem, even in principle. The reason is that the scale of supersymmetry breaking is much too high. It does, however, solve the Higgs mass or electroweak hierarchy problem in principle. The fact that this solution does not work (at least not very well) in practice is due to fairly recent data, especially from the LHC. Nevertheless, it will be important for us to study SUSY in general and to a certain extent the SUSY version of the Standard Model. The reasons are twofold. First, as noted, SUSY is an excellent example for how things could work out well at the cutoff scale Λ such that apparent fine tunings are at least mitigated. Second, if one wants quantum gravity divergences to also be tamed at the cutoff scale, SUSY is not enough and string theory is required. But the relation of the latter to real-world physics relies (at least in the best understood cases) on SUSY, which we hence have to understand at least at an introductory level.

2 Supersymmetry and Supergravity

There are many motivations to learn about SUSY. Let us give a few: SUSY is the only known symmetry relating fermions and bosons and may as such be a logical next step in the historical road towards unification in fundamental physics. String theory is the best-understood model of quantum gravity (or indeed the true theory underlying quantum gravity) and its stable versions all rely on SUSY (in 2d and in 10d). The only controlled roads from 10d strings to the 4d Standard Model involve 4d SUSY theories as an intermediate step. SUSY can resolve the hierarchy problem at the scale where it becomes manifest. (If it had been discovered at the electroweak scale, we could have found ourselves in a world without the hierarchy problem.) Even if that happens at, say, 10 TeV, the tuning would be much less severe than without SUSY. Finally, SUSY is a central tool in formal field theory research since SUSY theories usually involve many cancellations at the loop-level making them much better controlled. For example the, best-understood example of AdS/CFT involves an $\mathcal{N} = 4$ super-Yang-Mills (SYM) theory.

The structure and notation of what follows will be strongly influenced by the classic text [35], but there are many other useful books [36–39].

2.1 SUSY algebra and superspace

Recall the Poincare algebra

$$
\begin{aligned}
[P_\mu, P_\nu] &= 0 && (2.1)\\
[M_{\mu\nu}, P_\rho] &= i\eta_{\mu\rho}P_\nu - i\eta_{\nu\rho}P_\mu && (2.2)\\
[M_{\mu\nu}, M_{\rho\sigma}] &= i\eta_{\mu\rho}M_{\nu\sigma} + \cdots + \cdots && (2.3)
\end{aligned}
$$

as the symmetry algebra of $\mathbb{R}^{1,3}$. This algebra can be represntred by differential operators acting on functions on $\mathbb{R}^{1,3}$, e.g.

$$P_\mu = -i\partial_\mu \qquad \left(\partial_\mu = \frac{\partial}{\partial x^\mu}\right). \tag{2.4}$$

Indeed, these operators generate translations according to

$$\exp[i\epsilon^\mu P_\mu]\, f(x) = f(x) + \epsilon^\mu \partial_\mu f(x) + \cdots = f(x+\epsilon)\,. \tag{2.5}$$

Finite rotations in $\mathbb{R}^{1,3}$ are analogously generated by $M_{\mu\nu}$.

Any relativistic QFT has the above symmetry, but it may have additional ('internal') symmeties acting on the fields. Examples are a shift symmtry $\phi \to \phi + \epsilon$ or rotations in field space $\Phi \to \exp(i\epsilon_a T_a)\Phi$ with $\Phi \in \mathbb{C}^N$ and T_a the $SU(N)$ generators. Here 'additional' means that the full symmetry algebra is the direct sum of Poincare and internal algebra. The **Coleman-Mandula theorem** [33] claims that such a direct sum structure is the only possibility for how the Poincare-Algebra can be extended to a larger symmetry of a QFT (more precisely, of the S-matrix).

This theorem can be avoided if one generalizes the definition of a symmetry by a Lie algebra: One replaces the latter by a so-called a Lie superalgebra. Moreover, the resulting extension of the Poincare algebra is (essentially) unqiue and is called the supersymmetry algebra. This uniqueness is the statement of the **Haag-Lopuszanski-Sohnius theorem** [34].

We will not demonstrate uniqueness but only present the result of the analysis: The new generators to be added are (Weyl) spinors Q_α and the crucial new algebra relations are

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2(\sigma^\mu)_{\alpha\dot\alpha}P_\mu\,, \qquad \{Q_\alpha, Q_\beta\} = 0\,, \qquad \{\overline{Q}_{\dot\alpha}, \overline{Q}_{\dot\beta}\} = 0\,. \tag{2.6}$$

The main novelty is that for these generators one does not provide commutators but anticommuntators, hence we are now dealing with a Lie superalgebra.

The object $(\sigma^\mu)_{\alpha\dot\alpha}$ is defined as

$$\sigma^\mu = (-\mathbb{1}, \sigma^1, \sigma^2, \sigma^3) \tag{2.7}$$

and is an invariant tensor of $SL(2,\mathbb{C})$ just like $(\gamma^\mu)_{ab}$ is an invariant tensor of $SO(1,3)$. In fact, these two statements are of course related since the Lie algebras are the same and, roughly speaking, γ consists of two blocks of σs. We will return to this in more technical detail. One can avoid σs and Weyl spinors and formulate everything using left-handed 4-spinors, but Weyl spinors are very convenient in this context.

Relations between two bosonic generators remain commutators and relations between the new fermionic and the old bosonic commutators are also formulated in terms of commutators:

$$[P_\mu, Q_\alpha] = 0\,, \qquad [M_{\mu\nu}, Q_\alpha] = i(\sigma_{\mu\nu})_\alpha{}^\beta Q_\beta\,, \tag{2.8}$$

where

$$\sigma_{\mu\nu} \equiv -\frac{1}{4}(\sigma_\mu\overline{\sigma}_\nu - \sigma_\nu\overline{\sigma}_\mu) \qquad \text{and} \qquad (\overline{\sigma}_\mu)^{\dot{\alpha}\alpha} \equiv \epsilon^{\dot{\alpha}\dot{\beta}}\epsilon^{\alpha\beta}(\sigma^*_\mu)_{\dot{\beta}\beta}\,. \tag{2.9}$$

We will often use an overline instead of the star (or dagger) for complex conjugation (or the adjoint operator). The overline on $\overline{\sigma}$ does not specify whether upper or lower indices are assumed. Indices can be raised or lowered using the ϵ tensor. Given that we need the lower-upper index version of $(\sigma_{\mu\nu})_\alpha{}^\beta$ in (2.8), the expression on the l.h. side of (2.9) should be read as defining precisely this version. Hence, it involves a upper-upper index version of $\overline{\sigma}$, which is provided on the r.h. side of (2.9).

The full SUSY algebra is defined by (2.1)-(2.3) together with (2.6) and (2.8). Thus, we see that it consists of the Poincare algebra, the Q anticommutators, and the claim that the Qs transform under the Poincare algebra as space-time independent spinors. It may at this point also be useful to say more formally what a Lie superalgebra is: It is a vector space with a $\mathbb{Z}_2$ grading (it splits in an even and odd part) and with a binary operation that obeys the rules even$\times$even$\to$even, even$\times$odd$\to$odd and odd$\times$odd$\to$even. Furthermore, there are rules concerning the symmetries of these operations and Jacobi-like identities. These are, however, automatically fulfilled if the operations are explicitly realized through (anti)commutators, as in our case.

Next we want to represent this algebra on a larger space, called superspace. Its coordinates are

$$x^\mu\ (\mu = 0\cdots 3) \qquad \text{and} \qquad \theta^\alpha\ (\alpha = 1, 2)\,, \tag{2.10}$$

the latter being fermionic (Grassmann variables) and forming a Weyl spinor. The key relations for our purposes are

$$(\theta^\alpha)^* = \overline{\theta}^{\dot{\alpha}}\,, \qquad \{\theta^\alpha, \theta^\beta\} = 0 \quad \text{and} \quad \text{h.c.}\,, \qquad \{\theta^\alpha, \overline{\theta}^{\dot{\alpha}}\} = 0 \tag{2.11}$$

or, more explicitly,

$$(\theta^1)^2 = (\theta^2)^2 = 0\ , \qquad \theta^1\theta^2 = -\theta^2\theta^1\ , \qquad \text{etc.} \tag{2.12}$$

One also defines partial derivatives

$$\partial_\alpha = \frac{\partial}{\partial\theta^\alpha} \qquad \overline{\partial}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} \tag{2.13}$$

together with the obvious rules

$$\partial_\alpha\theta^\beta = \delta_\alpha{}^\beta\ , \qquad \overline{\partial}_{\dot{\alpha}}\overline{\theta}^{\dot{\beta}} = \delta_{\dot{\alpha}}{}^{\dot{\beta}}\ , \qquad \partial_\alpha\overline{\theta}^{\dot{\beta}} = 0 \qquad \overline{\partial}_{\dot{\alpha}}\theta^\beta = 0\,. \tag{2.14}$$

The reader should check that, as a result of the anticommutation relations for the θs, the ∂s also anticommute.

The space parameterized by the x^μ and θ^α is called superspace, in this case $\mathbb{R}^{4|4}$, with 4 bosonic and 4 real fermionic (or two complex fermionic) dimensions. Intuitively, one may want to think of $\mathbb{R}^4$ not as a set of points but, equivalently, as the algebra of functions on $\mathbb{R}^4$: $\{1, x^\mu, x^\mu x^\nu, \cdots\}$. The generalization to superspace is then obvious: One simply thinks of the algebra of functions including θs, i.e. $\{1, x^\mu, \theta^\alpha, x^\mu x^\nu, x^\mu\theta^\alpha, \cdots\}$.

Next, we naturally expect that a symmetry of this enlarged space will involve some analogue of the familiar generators of translations, i.e. $Q_\alpha \sim \partial_\alpha + \cdots$. The ellipsis stands for extra terms which must come in to ensure that Qs anticommute to give the Ps. The correct formulae turn out to be

$$Q_\alpha = \partial_\alpha - i(\sigma^\mu)_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu\,, \qquad \overline{Q}_{\dot\alpha} = -\overline{\partial}_{\dot\alpha} + i\theta^\alpha(\sigma^\mu)_{\alpha\dot\alpha}\partial_\mu\,. \tag{2.15}$$

It is a straightforward but very important excercise to derive the essential part of the SUSY algebra from this:

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2i(\sigma^\mu)_{\alpha\dot\alpha}\partial_\mu\,, \qquad \{Q_\alpha, Q_\beta\} = 0\,, \qquad \{\overline{Q}_{\dot\alpha}, \overline{Q}_{\dot\alpha}\} = 0\,. \tag{2.16}$$

Let us pause for a small, technical comment: The reader will have noticed that, with the standard identification $P_\mu = -i\partial_\mu$ (recall that we are using a mostly-plus metric), the algebras of (2.6) and (2.16) differ by a sign. This is nothing deep but merely a result of two different ways of defining the operators Q and P. On the one hand, one may think of them as acting on functions. On the other hand, as acting on coordinates. To make this clear, one may consider the relation

$$(\hat{A}\hat{B}f)(x) = (\hat{B}f)(Ax) = f(BAx) \tag{2.17}$$

between operators A, B acting on coordinates x and the corresponding operators $\hat{A}$, $\hat{B}$ acting on functions of x. It is immediately clear from the above that

$$[\hat{A}, \hat{B}] = \hat{C} \qquad \text{and} \qquad [A, B] = -C \tag{2.18}$$

follows if the operators C and $\hat{C}$ are related as explained above for A, B and $\hat{A}, \hat{B}$. Thus, the sign in the Lie algebra flips between the two different definitions.

2.2 Superfields

Now one builds a field theory on this enlarged space. A (complex) general superfield is a function

$$\begin{aligned} F(x,\theta,\overline{\theta}) \quad = \quad & f(x) + \theta\phi(x) + \overline{\theta}\overline{\chi}(x) + \theta^2 m(x) + \overline{\theta}^2 n(x) + \theta\sigma^\mu\overline{\theta}v_\mu(x) \\ & + \theta^2\overline{\theta}\overline{\lambda}(x) + \overline{\theta}^2\theta\psi(x) + \theta^2\overline{\theta}^2 d(x)\,. \end{aligned} \tag{2.19}$$

Here the r.h. side is a Taylor expansion of the l.h. side where, however, all higher terms vanish. Here ϕ, χ, λ and ψ are Weyl spinors, anticommuting among each other and with the θs.

We have started to use a very convenient shorthand notation for the product of Weyl spinors, for example

$$\theta\phi \equiv \theta^\alpha\phi_\alpha = \epsilon^{\alpha\beta}\theta_\alpha\phi_\beta\ , \quad \text{and analogously} \quad \theta^2 = \theta^\alpha\theta_\alpha\,. \tag{2.20}$$

It is an essential part of this convention that suppressed undotted indices are always summed from upper-left to lower-right. For dotted indices, the rule is inverse:

$$\overline{\theta}\overline{\chi} \equiv \overline{\theta}_{\dot\alpha}\overline{\chi}^{\dot\alpha}\,. \tag{2.21}$$

This convention goes together with certain ϵ tensor conventions:

$$\epsilon_{\alpha\beta} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\,, \qquad \epsilon^{\alpha\beta} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\,, \qquad \text{with} \qquad \epsilon_{\alpha\beta}\epsilon^{\beta\gamma} = \delta_\alpha{}^\gamma\,. \tag{2.22}$$

With this contraction one has, in spite of the anticommutation relations,

$$\psi\chi = \chi\psi\,, \tag{2.23}$$

as the reader should check.

It goes without saying that the Poincare algebra acts on superfields in the usual way, e.g.

$$\delta_\epsilon F = i\epsilon^\mu P_\mu F = \epsilon^\mu \partial_\mu F\,. \tag{2.24}$$

By analogy, we define the **SUSY transformation**

$$\delta_\xi F = (\xi Q + \overline{\xi Q})F = [\,(\xi\partial - i\xi\sigma^\mu\overline{\theta}\partial_\mu) \,+\, \text{h.c.}\,]F\,. \tag{2.25}$$

Here by 'h.c.' we mean the application of a formal $*$-operation on the algebra of functions and differential operators. In essence, this is just complex conjugation and its obvious extension to the θs. A crucial exception is the rule

$$(\partial_\alpha)^* = -\overline{\partial}_{\dot\alpha}\,, \tag{2.26}$$

which is required by consistency. The reader should check this by carefully thinking about the possible ways to evaluate $(\partial_\alpha\theta^\beta)^*$.

Returning to our SUSY transformations, we note that the superfield F is of course just an abstract concept useful for the defining 'component' fields f, ϕ_α etc., which are conventional quantum fields. Thus, after calculating $\delta_\xi F$, we expand it in a Taylor series and define $\delta_\xi f$, $\delta_\xi\phi$, etc. as the coefficients of the various terms with growing powers of θ:

$$\delta_\xi F = \delta_\xi f + \theta^\alpha(\delta_\xi\phi)_\alpha + \cdots\,. \tag{2.27}$$

This defines the SUSY transformation of the component fields.

2.3 Chiral superfields

The general superfield is too large to be practically useful and it does indeed correspond to a reducible representation of the SUSY algebra. Simpler superfields exist and are sufficient to write down the most general SUSY lagrangian.

To define the chiral superfield, it is useful to first introduce SUSY-covariant derivatives (in a way very similar to the Qs):

$$D_\alpha = \partial_\alpha + i(\sigma^\mu)_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu\,, \qquad \overline{D}_{\dot\alpha} = -\overline{\partial}_{\dot\alpha} - i\theta^\alpha(\sigma^\mu)_{\alpha\dot\alpha}\partial_\mu\,. \tag{2.28}$$

They obey

$$\{D_\alpha, \overline{D}_{\dot\alpha}\} = -2i(\sigma^\mu)_{\alpha\dot\alpha}\partial_\mu\,, \qquad \{D_\alpha, D_\beta\} = 0\,, \qquad \{\overline{D}_{\dot\alpha}, \overline{D}_{\dot\alpha}\} = 0 \tag{2.29}$$

and, crucially, any D or $\overline{D}$ anticommutes with any Q or $\overline{Q}$,

$$\{D_\alpha, \overline{Q}_{\dot\alpha}\} = 0 \qquad \text{etc.} \tag{2.30}$$

This last feature implies that

$$\overline{D}_{\dot\alpha} F = 0 \qquad \Rightarrow \qquad \overline{D}_{\dot\alpha} \delta_\xi F = 0\,. \tag{2.31}$$

In other words, superfields fullfilling the condition $\overline{D}_{\dot\alpha} F = 0$ form a subrepresentation of the Lie superalgebra representation provided by general superfields. We call them **chiral superfields**.

It can be shown that chiral superfields can be written as

$$\Phi = \Phi(y, \theta) \qquad \text{with} \qquad y^\mu = x^\mu + i\theta\sigma^\mu\overline{\theta} \tag{2.32}$$

and expanded according to

$$\Phi = A(y) + \sqrt{2}\theta\psi(y) + \theta^2 F(y)\,. \tag{2.33}$$

As explained above for the general superfield, one obtains the SUSY transformations of the component fields by expanding $\delta_\xi\Phi$ in the same way as Φ. The result reads

$$\begin{aligned} \delta_\xi A &= \sqrt{2}\psi\xi \\ \delta_\xi \psi &= i\sqrt{2}\sigma^\mu\overline{\xi}\partial_\mu A + \sqrt{2}\xi F \\ \delta_\xi F &= i\sqrt{2}\overline{\xi}\overline{\sigma}^\mu\partial_\mu\psi\,. \end{aligned} \tag{2.34}$$

We note that one can analogously define antichiral superfields, $D_\alpha\overline{\Phi} = 0$, and that the conjugate of a chiral superfield is antichiral.

2.4 SUSY-invariant lagrangians

We state without proof that the most general such lagrangian, at the 2-derivative-level and built from chiral superfields $\{\Phi_1, \cdots, \Phi_n\}$ only, reads

$$\mathcal{L} = K(\Phi_i, \overline{\Phi}_{\bar\imath})\Big|_{\theta^2\overline{\theta}^2} + W(\Phi_i)\Big|_{\theta^2} + \text{h.c.}\,. \tag{2.35}$$

Here K is a real function of a set of complex variables Φ_i. With Φ_i being chiral superfields, K becomes a general superfield. It is not chiral since both Φ_i and $\overline{\Phi}_{\bar\imath}$ are involved. The first term in $\mathcal{L}$ is the projection of the general superfield K on its highest component, i.e., it is the analogue of the function $d(x)$ appearing in the Taylor expansion (2.19).

The function K is called the Kahler potential (for those who know this term from complex geometry: the relevance in the present context will become clear momentarily). The expression $K\Big|_{\theta^2\overline{\theta}^2}$, viewed as part of the lagrangian, is called the D-term. This name comes simply from the traditional use of the variable $d(x)$ for the highest component. The key point in this non-trivial way of writing a lagrangian is, of course, its required invariance under SUSY transformations. For this, we need to recall that the commutator of Q and $\overline{Q}$ is P. Hence the mass dimension of Q is 1/2. Since Q involves $\partial/\partial\theta$, the mass dimension of θ is $-1/2$ (one may think of it very vaguely as the square root of x). Thus, in the Taylor expansion of superfields in powers of θ the mass dimensions of components grow. As a result, due also to the linear nature of SUSY transformations, the highest component can not transform into any other component – there simply is no component with a suitably high mass dimensions. The only way it can transform is into a *derivative* of another component. Thus, the first term of the above lagrangian is invariant up to total derivatives, as one would have hoped.

Similarly, W is called the superpotential and it is an analytic (or holomorphic) function of the Φ_i. This makes W a chiral superfield. In its Taylor expansion in θ, with the coefficients being functions of y, the highest component is tradtionally called F (cf. (2.33)). Hence the corresponding two terms in (2.35) are sometimes called F terms. To be very concrete, to get these terms one expands the chiral superfield $W(\Phi_i)$ in θ (with the coefficients being functions of y), extracts the coefficient of θ^2, and replaces y by x. The result, together with its hermitian conjugate, is the F-term lagrangian. It is SUSY invariant up to a total derivative for the same reason as explained in the case of the D term.

An equivalent way of writing this lagrangian is as

$$\mathcal{L} = \int d^2\theta\, d^2\overline{\theta}\, K(\Phi_i, \overline{\Phi}_{\bar{\imath}}) + \int d^2\theta\, W(\Phi_i) + \text{h.c.}\,. \tag{2.36}$$

Using standard integration rules for Grassmann variables,

$$\int d\theta_1\, \theta_1 = 1 \qquad \text{and} \qquad \int d\theta_1\, 1 = 0\,, \tag{2.37}$$

and the analogous formulae for θ_2, one can easily check that the integral formulation is equivalent to the projection formulation of $\mathcal{L}$. The SUSY invariance is particularly easily seen in the integral formulation: The SUSY generator Q is a linear combination of x derivatives and θ derivatives. The x derivative of any lagrangian is, by definition, a total derivative and thus leaves the action invariant. The θ derivative of any expression in θ integrates to zero,

$$\int d\theta_1\, \frac{\partial}{\partial\theta_1}\Big(\cdots\Big) = 0\,, \tag{2.38}$$

as one can easily convince oneself. Thus, any action which is an integral over the full superspace is invariant. Similarly, any action built as the integral of an expression in θ (not $\overline{\theta}$) and integrated over half the superspace is invariant. (Here it is important to note that we can replace y with x by appealing to the Taylor expansion and the irrelevance of total derivatives.)

2.5 Wess-Zumino-type models

The possibly simplest interesting SUSY model is the Wess Zumino model. It is defined by

$$K = \overline{\Phi}\Phi\ , \qquad W = \frac{m}{2}\Phi^2 + \frac{\lambda}{3}\Phi^3\,. \tag{2.39}$$

A straightforward explicit calculation according to the rules above gives the following component form of the lagrangian:

$$\mathcal{L} = -|\partial A|^2 - i\overline{\psi}\overline{\sigma}^\mu\partial_\mu\psi + \Big(-\frac{m}{2}\psi^2 + \lambda\psi^2 A\Big) + \text{h.c.} + (mA + \lambda A^2)F + \text{h.c.} + |F|^2\,. \tag{2.40}$$

Since F has no kinetic term (and thus does not propagate) we can integrate it out by purely algebraic operations and without any approximation. Such fields are called auxiliary fields. The equation of motion for F is

$$F = -m\overline{A} - \lambda\overline{A}^2\,, \tag{2.41}$$

and inserting this into the original lagrangian gives

$$\mathcal{L} = -|\partial A|^2 - i\overline{\psi}\overline{\sigma}^\mu\partial_\mu\psi + \Big(-\frac{m}{2}\psi^2 + \lambda\psi^2 A\Big) + \text{h.c.} - V(A,\overline{A})\,, \tag{2.42}$$

with the scalar potential (or F term potential)

$$V(A,\overline{A}) = |F|^2 = |mA + \lambda A^2|^2\,. \tag{2.43}$$

This is easily generalized to (non-renormalizable and multi-field) models of the type

$$K = K(\Phi_i, \overline{\Phi}_{\bar{\imath}})\;, \qquad W = W(\Phi_i)\,. \tag{2.44}$$

We only display the purely bosonic part of the resulting component lagrangian. More details are given in the problems. The auxiliary fields have already been integrated out:

$$\mathcal{L} = K_{i\bar{\jmath}}(A_k, \overline{A}_{\bar{k}})\,(\partial A_i)(\partial\overline{A}_{\bar{\jmath}}) + K^{i\bar{\jmath}}(A_k, \overline{A}_{\bar{k}})\,(\partial_i W(A_k))(\overline{\partial}_{\bar{\jmath}}\overline{W}(\overline{A}_{\bar{k}})) + \cdots\,. \tag{2.45}$$

Here

$$K_{i\bar{\jmath}} = \partial_i\overline{\partial}_{\bar{\jmath}}K \qquad \text{and} \qquad K_{i\bar{\jmath}}K^{k\bar{\jmath}} = \delta_i{}^k\,, \tag{2.46}$$

in other words, indices denote partial derivatives and the upper-index matrix is defined as the inverse.

We note that the scalar components A_i parametrize a complex manifold (as in so-called sigma-models) and, in supersymmetry, the metric on this field space is the Kahler metric $K_{i\bar{\jmath}}$, defined with the help of the Kahler potential K. The superpotential W is locally a holomorphic function on this manifold; globally it is a section in an appropriate complex line bundle.

2.6 Real Superfields

We have to discuss real superfields, another subrepresentation contained in that of the general superfield, since they are needed to describe gauge theories. But we will be very brief since, conceptually, the procedure is similar to that used in the chiral superfield case.

A real superfield $V = V(x,\theta,\overline{\theta})$ is defined by the condition $V = \overline{V}$. It can be Taylor expanded in θ and $\overline{\theta}$. We will build lagrangians which are invariant under the SUSY gauge transformation

$$2V \to 2V + \Lambda + \overline{\Lambda}\,, \tag{2.47}$$

with Λ a chiral superfield. Using this transformation, V can be brought to a form where certain components vanish (the so-called Wess-Zumino gauge):

$$V = -\theta\sigma^\mu\overline{\theta}A_\mu + i\theta^2\,\overline{\theta\lambda} - i\overline{\theta}^2\,\theta\lambda + \frac{1}{2}\theta^2\overline{\theta}^2 D\,. \tag{2.48}$$

From V, one can constuct the so-called field-strength superfield

$$W_\alpha = -\frac{1}{4}\overline{D}^2 D_\alpha V\,, \tag{2.49}$$

which can be shown to be chiral and gauge invariant. Its name is justified since it does indeed contain the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ in one of its components:

$$W = i\lambda(y) + [D(y) + i\sigma^{\mu\nu}F_{\mu\nu}(y)]\cdot\theta + \theta^2\sigma^\mu\partial_\mu\overline{\lambda}(y)\,. \tag{2.50}$$

One can show that SUSY gauge transformations contain standard gauge transformations as a subset. Hence, it is natural to look for SUSY-invariant and SUSY-gauge-invariant lagrangians. At the 2-derivative level, the unique option is

$$\mathcal{L} = \frac{1}{4g^2}\left(W^\alpha W_\alpha\Big|_{\theta^2} + \overline{W}_{\dot\alpha}\overline{W}^{\dot\alpha}\Big|_{\overline{\theta}^2}\right) = \frac{1}{g^2}\left\{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\overline{\lambda}\overline{\sigma}^\mu\partial_\mu\lambda + \frac{1}{2}D^2\right\}\,, \tag{2.51}$$

where λ is the gaugino and D a real auxiliary field.

It is straightforward to extend this to the non-abelian case, where V and W are matrix-valued superfields taking values in the Lie Algebra of the gauge group. Let us write the corresponding lagrangian including also a charged matter superfield Φ, to be thought of as column-vector in some appropriate representation. We have

$$\mathcal{L} = \frac{1}{2g^2}\,\mathrm{tr}\left(W^2\Big|_{\theta^2} + \text{h.c.}\right) + \Phi^\dagger e^{2V}\Phi\Big|_{\theta^2\overline{\theta}^2} + \mathcal{W}(\Phi)\Big|_{\theta^2} + \text{h.c.}\,. \tag{2.52}$$

Here e^{2V} has to be taken in the representation of Φ and $\Phi^\dagger$ has to be interpreted as a row vector. This lagrangian is invariant under the non-abelian super gauge transformations

$$e^{2V} \to e^{\Lambda^\dagger}e^{2V}e^{\Lambda}\,, \qquad \Phi \to e^{-\Lambda}\Phi\,. \tag{2.53}$$

One very frequently uses the naming conventions for components

$$\Phi = \{\Phi, \psi, F\}\,, \qquad V = \{A_\mu, \lambda, D\}\,. \tag{2.54}$$

It is a slight abuse of notation to denote the scalar matter component by the same name as the superfield, but this convention is widespread and it is usually clear from the content which meaning is intended. With these conventions, the component form of the lagragian reads

$$\begin{aligned}\mathcal{L} \;=\; & \frac{1}{g^2}\,\mathrm{tr}\left\{-\frac{1}{2}F_{\mu\nu}F^{\mu\nu} - 2i\overline{\lambda}\overline{\sigma}^\mu D_\mu\lambda + D^2\right\} \\ & -|D_\mu\Phi|^2 - i\overline{\psi}\overline{\sigma}^\mu D_\mu\psi + |F|^2 + i\sqrt{2}\left(\Phi^\dagger\lambda\psi - \overline{\psi}\overline{\lambda}\Phi\right) + \Phi^\dagger D\Phi\,.\end{aligned} \tag{2.55}$$

This lagrangian is called off-shell since it is SUSY invariant without using the equations of motion. Integrating out the auxiliary field, one arrives at the on-shell lagrangian. Concerning F, this step is trivial in this simple example: F is just set to zero. By contrast, integrating out $D = D_aT_a$ induces a quartic term in the scalar fields, the so-called D-term potential.

2.7 SUSY breaking

We have so far defined the spinor Q as a differential operator on superspace. Hence, it is an operator on the space of superfields, hence an operator transforming different component fields into each other. After quantization, we will thus be able to define a corresponding operator Q on the Hilbert space. This operator will mix bosons and fermions and, since

$$[Q_\alpha, P_\mu P^\mu] = 0\,, \tag{2.56}$$

this implies that the mass of fermions and bosons (in the same superfield or multiplet) is the same. Thus, to be relevant for the real world, supersymmetry must be spontaneously broken. In other words, while the action should be supersymmetric, the vacuum should not be invariant.

At the perturbative level, this simply means that the lowest-energy field configuration should not be invariant under SUSY. Thus, in the context of chiral superfields, the r.h. side of

$$\begin{aligned} \delta_\xi A &= \sqrt{2}\psi\xi \\ \delta_\xi \psi &= i\sqrt{2}\sigma^\mu\overline{\xi}\partial_\mu A + \sqrt{2}\xi F \\ \delta_\xi F &= i\sqrt{2}\overline{\xi}\overline{\sigma}^\mu\partial_\mu\psi \end{aligned} \tag{2.57}$$

should be non-zero. Maintaining Lorentz-invariance, this can only be achieved if $F \neq 0$ in the vacuum. This is called F-term breaking and the simplest lagrangian with this feature is

$$\mathcal{L} = \overline{\Phi}\Phi\Big|_{\theta^2\overline{\theta}^2} + c\Phi\Big|_{\theta^a} + \text{h.c.}\,. \tag{2.58}$$

The relevant terms in component form are

$$\mathcal{L} = \overline{F}F + cF + \text{h.c.} + \cdots\,, \tag{2.59}$$

which implies $\overline{F} = -c$ in the vacuum. However, while SUSY is formally broken, the theory is free, thus F does not couple to other fields and hence the spectrum remains supersymmetric.

This is easily remedied adding a higher-dimension operator,

$$\mathcal{L} = \left[\overline{\Phi}\Phi - \frac{1}{M^2}(\overline{\Phi}\Phi)^2\right]\Big|_{\theta^2\overline{\theta}^2} + c\Phi\Big|_{\theta^a} + \text{h.c.} \tag{2.60}$$

Now, ignoring fermions and derivative terms, the component lagrangain reads

$$\mathcal{L} = \overline{F}F - 4\overline{F}F\overline{\Phi}\Phi/M^2 + cF + \text{h.c.} + \cdots\,. \tag{2.61}$$

The vacuum is again at $\Phi = 0$ and $\overline{F} = -c$, but now this non-zero F introduces scalar masses and supersymmetry is broken in the spectrum of the theory.

We note that apparently simpler models which extend (2.58) by adding terms $\sim \Phi^2$ or $\sim \Phi^3$ to the superpotential do not work in our context. They reinstate a SUSY-preserving vacuum, which is obvious since in such models the linear term can be absorbed in a shift of Φ. In fact, the simplest renormalizable model with chiral superfields and spontaneous SUSY breaking is the **O'Raifeartaigh model** with lagrangian

$$\mathcal{L} = \sum_{i=1}^{3}\overline{\Phi}_i\Phi_i\Big|_{\theta^2\overline{\theta}^2} + \left[\Phi_1(m^2 + \lambda\Phi_3^2) + \mu\Phi_2\Phi_3\right]\Big|_{\theta^a} + \text{h.c.}\,. \tag{2.62}$$

It is easy to write down the F-term potential and minimize it to find spontaneous SUSY breaking. Sometimes the name O'Raifeartaigh model is used more generally for any model with F-term breaking.

A completely analogous story can be developed for real superfields, i.e. (abelian) gauge theories, where SUSY breaking is signalled by a non-zero VEV of the D-term. The simplest model realizing this is

$$\mathcal{L} = \frac{1}{2g^2} W^2 \Big|_{\theta^2} + 2\kappa V \Big|_{\theta^2\overline{\theta}^2}\,, \tag{2.63}$$

where the new term linear in V is known as Fayet-Ilopoulos or FI term. At the component level one finds

$$\mathcal{L} = \frac{1}{2g^2} D^2 + \kappa D \qquad \Rightarrow \qquad D = -\kappa g^2 \neq 0\,. \tag{2.64}$$

As before, the model needs to be enriched to see this formally present SUSY breaking in the spectrum. This can be achieved e.g. by adding two chiral superfields (to avoid anomalies) with charge ± 1 and mass m. One finds that the femions remain massless while the boson masses split according to $m_{1,2}^2 = m^2 \pm \kappa g^2$.

2.8 Supersymmetrizing the Standard Model

The Minimal Supersymmetric Standard Model or MSSM is obtained basically by promoting all fermions and scalars of the Standard Model to chiral superfields and all vectors to real superfields. The additional components introduced in this way are made heavy by an appropriate mechanism of SUSY breaking, to be discussed shortly. The only exception to this rule arises in the Higgs sector, where one now needs two different Higgs doublets and hence two corresponding superfields: H_u and H_d. The reason will become clear immediately. Of the many reviews of this wider subject we refer in particular to [17, 41, 42].

After these preliminaries, we give the set of chiral superfields,

$$\Phi_a = \{Q, U, D, L, E, H_u, H_d\}\,, \tag{2.65}$$

in complete analogy with (1.2) and with gauge charges as in (1.6). The lagrangian can be organized in three pieces. First,

$$\mathcal{L}_{gauge} = \sum_{i=1}^{3} \mathrm{tr}\left(\frac{1}{2g_i^2} (W_i)^\alpha (W_i)_\alpha \Big|_{\theta^2} + \text{h.c.} \right)\,, \tag{2.66}$$

with tr $\to 1/2$ in the $U(1)$ case. Second,

$$\mathcal{L}_K = \sum_{a=1}^{7} \Phi_a^\dagger e^{2V} \Phi_a \Big|_{\theta^2\overline{\theta}^2}\,, \tag{2.67}$$

where K stands for kinetic or Kahler potential term and where the superfield $V = V_1 + V_2 + V_3$ contains the three real superfields corresponding to three factors of G_{SM} and one must always use the representation appropriate for Φ_a. Third, we have the superpotential term

$$\mathcal{L}_W = (W_\mu + W_Y + W_e)\Big|_{\theta^2} + \text{h.c.} \tag{2.68}$$

with

$$W_\mu = \mu H_u H_d\,, \qquad W_Y = \lambda_u Q H_u U + \lambda_d Q H_d D + \lambda_e L H_d E\,, \tag{2.69}$$

and

$$W_e = aLH_u + bQLD + cUUD + dLLE\,. \tag{2.70}$$

The structure of $\mathcal{L}_{gauge}$ and $\mathcal{L}_K$ requires no further comments: They simply provide the necessary kinetic terms and gauge interactions. The Yukawa couplings come, together with new interactions, from W_Y. To give masses to all fermions, we were forced to introduce two Higgs fields. Indeed, holomorphicity forbids the appearance of $\tilde{\Phi}$ used in the up-type Yukawa in the non-supersymmetric Standard Model. Hence, a new Higgs multiplet H_u with opposite $U(1)$ charge has to be introduced. An independent reason for this second doublet is the need to cancel the $U(1)$ anomaly introduced by the fermion (the 'Higgsino') contained in H_d.

Finally, there are *extra* terms without a Standard Model analogue, allowed due to the enlarged field content. We have collected these terms in W_e but, since some of them induce proton decay and lepton number violation, we basically want to forbid them. We also note that we have limited ourselves to the renormalizable level – hence W is truncated at cubic order. To see that cubic terms correspond to marginal operators, recall that θ^2 has mass dimension -1. Hence, projection on the θ^2 component corresponds to raising the mass dimension by one unit. Thus, mass dimension 3 in W corresponds to mass dimension 4 in $\mathcal{L}$.

To forbid W_e, the concept of an R**-symmetry** (which is crucial in SUSY independently of phenomenology) is useful. To explain this concept, we define standard (global) $U(1)$ and $U(1)_R$ transformations of chiral superfields as follows:

$$U(1):\ \Phi(y,\theta) \to e^{im\epsilon}\Phi(y,\theta)\ , \qquad U(1)_R:\ \Phi(y,\theta) \to e^{in\epsilon}\Phi(y, e^{-i\epsilon}\theta)\,. \tag{2.71}$$

Here m and $n = R(\Phi)$ are the $U(1)$ and $U(1)_R$ charges of Φ respectively. It follows immediately that, and this is the crucial feature of an R symmetry, the components transform differently, depending on their mass dimension:

$$A \to e^{in\epsilon}A\,, \qquad \psi \to e^{i(n-1)\epsilon}\psi\,, \qquad F \to e^{i(n-2)\epsilon}F\,. \tag{2.72}$$

Invariance of the lagrangian requires $R(K) = 0$ and $R(W) = 2$. The former is clear since the projection on the $\theta^2\overline{\theta}^2$ component does not change the R-charge. By contrast, projection on the θ^2 component lowers the R-charge by 2.

For our purposes, the interesting assignment is:

$$\text{For}\ \ Q,U,D,L,E:\ \ R=1 \qquad\qquad \text{and for}\ \ H_u,H_d:\ \ R=0\,. \tag{2.73}$$

This restricts W to the Yukawa terms. However, this is too strong since it also forbids the so-called μ term $\mu H_u H_d$. But the latter is needed since even after SUSY breaking (see below) it is the only source for Higgsino masses. (Higgsinos are the - so far unobserved and hence heavy - fermionic partners of the Higgs).

A possible resolution is the breaking of $U(1)_R$ to its $\mathbb{Z}_2$ subgroup, defined by restricting $\{\, e^{i\epsilon}\,\}$ to the two elements with $\epsilon = 0$ and $\epsilon = \pi$. After this breaking to $\mathbb{Z}_2$, R-charges are identified modulo 2. Indeed, superfields with R-charge m and $m+2$ now transform identically. In particular the selection rule $R(W) = 2$ for superpotential terms is modified to $R(W) = 2 \mod 2$. In other words, one now only demands $R(W) \in 2\mathbb{N}$. As a result, the μ term is allowed while all terms in W_e are still forbidden. Moreover, the transformation rules of the Standard Model fields and their superpartners under this so-called $\mathbb{Z}_2$ R**-parity** are

$$\textbf{Even:} \qquad \text{Higgs scalars, fermions, gauge bosons} \tag{2.74}$$

$$\textbf{Odd:} \qquad \text{Higgsinos, sfermions, gauginos}\,. \tag{2.75}$$

This implies that any of the so-called superpartners can not decay into a combination of Standard Model particles. Hence the lightest superpartner (the **lightest supersymmetric particle or LSP**) is absolutely stable and provides a natural dark matter candidate. Unfortunately, with growing LHC-bounds on its mass the abundance predicted from its so-called freeze-out in early cosmology tends to become too high, calling for extensions of the simplest settings.

2.9 Supersymmetric and SUSY breaking masses and non-renormalization

The simplest way to make the above construction realistic is to add mass terms to the supersymmetric Standard Model such that all remaining superpartners become sufficiently heavy. (Recall that the Higgsino can be made heavy by a sufficiently large μ term.) While technically correct, such an approach of explicit SUSY breaking is not very satisfying or illuminating concerning the resolution of the hierarchy problem.

Hence, we will introduce somewhat more structure and try to arrive at the MSSM using spontaneous SUSY breaking. Specifically, we introduce a **hidden sector** in which SUSY is broken spontaneously. [8] It will then be communicated to the Standard Model by higher-dimension operators. To illustrate this structure, we start with the toy-model lagrangian

$$\mathcal{L} = \left(\overline{S}S - c_1(\overline{S}S)^2\right)\Big|_{\theta^2\overline{\theta}^2} + c_2 S\Big|_{\theta^2} + \text{h.c.} + \overline{\Phi}\Phi\Big|_{\theta^2\overline{\theta}^2} + \frac{1}{M^2}\overline{\Phi}\Phi\overline{S}S\Big|_{\theta^2\overline{\theta}^2}\,. \tag{2.76}$$

We recognize a model with a chiral superfield S and with spontaneous SUSY breaking ($F_S \neq 0$). In addition, we have free and massless chiral superfield Φ. The latter represents the Standard Model or, more specifically, its Higgs sector. Finally, the last term is a higher-dimension operator, suppressed by a large mass scale M, coupling the two sectors. All we need to know about the hidden sector is that $S = 0$ and $F_S \neq 0$ in the vacuum. Inserting this in our lagrangian and focussing on the Φ-sector only, we have

$$\mathcal{L} \supset \overline{\Phi}\Phi\Big|_{\theta^2\overline{\theta}^2} + \frac{1}{M^2}\overline{A}_\Phi A_\Phi \overline{F}_S F_S\,, \tag{2.77}$$

where we also ignored a quartic fermionic term arising from the superfield higher-dimension operator. We see that the result is equivalent to just having added a ('soft') SUSY-breaking scalar mass term to the Φ sector

$$\mathcal{L} \supset m_{soft}^2 |A_\Phi|^2 \qquad \text{with} \qquad m_{soft}^2 = |F_S|^2/M^2\,. \tag{2.78}$$

[8]In principle, one may imagine situations where SUSY is broken spontaneuously in the supersymmetrized Standard Model, without introducing any additonal fields. However, it turns out that this does not work in practice, taking into account experimental constraints on masses and the phenomenologically required gauge symmetry breaking.

Crucially, in our approach we see right away that this term is radiatively stable - it is secretly a higher-dimension operator and does as such not receive power-divergent loop corrections. This explains the name 'soft'. In fact, the two sectors decouple completely as $M \to \infty$, making it clear that the coupling operator can only renormalize proportionally to itself. (We see here another possibility, in addition to symmetries, why a certain coefficient in the lagrangian may be zero in a natural way: In its absence, the model becomes the sum of two completely independet theories.)

Our point about the mass term not being quadratically divergent may appear trivial - after all the Φ sector itself is free theory, so of course nothing renormalizes. However, it is immediate to enrich our model by e.g. $W(\Phi) \sim \Phi^3$, leading to quartic self interactions. Alternatively, Φ may be charged under some gauge group, like the Higgs in the Standard Model is. Nothing in our argument changes: The operator $\sim 1/M^2$ inducing the mass can not have power-divergences.

However, one could clearly add a term $W(\Phi) \sim m_{SUSY}\Phi^2$ to our action, in other words, a supersymmetric mass term. We have to be sure that interactions in the Φ sector will not, if such a term is absent in the beginning, induce it through loop corrections. This, as it turns out, is in fact the main point where SUSY saves us: **The superpotential does not renormalize**. This so called non-renormalization theorem is, at least at a superficial level and in our simple model, easy to understand [40]:

Indeed, consider the WZ model with tree level superpotential

$$W = \frac{m}{2}\Phi^2 + \frac{\lambda}{3}\Phi^3 \,. \tag{2.79}$$

Introduce a $U(1)$ and $U(1)_R$ symmetry under which Φ has charges $(1,1)$. Clearly, this is respected by our canonical Kahler potential K, but the superpotential breaks both symmetries. One can interpret this breaking as being due to non-zero VEVs of superfields m and λ, the scalar components of which have acquired non-zero VEVs. For this interpretation to work, one needs to assign to m the charges $(-2,0)$ and to λ the charges $(-3,-1)$. Now, assuming that perturbative loop corrections break neither these $U(1)$ symmetries nor SUSY, we expect that the effective superpotential (relevant for the Wilsonian effective action) will still respect the two $U(1)$ symmetries. Using holomorphicity and the fact that each term in W must have charges $(0,2)$, we have

$$W_{\text{eff}} = \sum_{ijk} c_{ijk} m^i \lambda^j \Phi^k = m\Phi^2\, f(\lambda\Phi/m) \,. \tag{2.80}$$

In the second step, we used the fact that, under the symmetry constraints, the triple sum collapses to a single sum, which can then be viewed as a power series in $(\lambda\Phi/m)$. This last combination of fields can appear to any power, since both its $U(1)$ and $U(1)_R$ charge vanish.

Now, the constant and linear term in f correspond to the terms already present at tree level - their values are $1/2$ and $1/3$ by assumption. We see that higher terms in λ, which may in principle arise from loop corrections, always come with higher powers of Φ and hence do not affect mass and trilinear coupling. Moreover, it is easy to convince oneself that such higher terms in λ, as derived from (2.80), correspond precisely to terms following from tree-level exchange of Φ. But such tree-level effects should not be included in W_{eff}. They appear in the calculation of observable if one uses only tree-level expression for W together with the standard Feynman rules. Including them in W_{eff} would lead to double counting. Now, compared to tree-level effects, loop effects always have a higher power of λ (given a certain number of external legs). Hence such loop effects are not described by the higher-λ terms in f. As a result, we learn that $W_{\rm eff} = W$ and no loop corrections arise.

In summary, we have learned that the structure of (2.76), supplemented by a superpotential of type (2.79), is radiatively stable. In particular, the supersymmetric and supersymmetry-breaking mass terms can both be chosen small compared to the cutoff scale and are not subject to power-like divergences.

2.10 The Minimal Supersymmetric Standard Model (MSSM)

With this, it is straightforward to introduce SUSY breaking by a spurion superfield S into the SUSY Standard Model. Without aiming at completeness, we give three types of higher-dimension terms which are sufficient to generate all essential SUSY breaking effects:

$$
\begin{aligned}
\mathcal{L}_1 &= \frac{1}{M^2}\overline{Q}Q\overline{S}S\Big|_{\theta^2\bar\theta^2} \quad , \quad & \mathcal{L}_2 = \frac{1}{M^2}Q^2\overline{S}S\Big|_{\theta^2\bar\theta^2} \\
\mathcal{L}_3 &= \frac{1}{M}Q^3S\Big|_{\theta^2} + \text{h.c.} \quad , \quad & \mathcal{L}_4 = \frac{1}{M}W^\alpha W_\alpha S\Big|_{\theta^2} + \text{h.c.}
\end{aligned}
\tag{2.81}
$$

Here Q stands for generic Standard Model chiral superfields. The different factors of Q in one term may also be replaced by different Standard Model fields, e.g. $Q^3 \to QH_uU$.

The effects of these different terms are easy to read off. For example,

$$
\mathcal{L}_1 \supset \frac{|F_S|^2}{M^2}|A_Q|^2 \equiv M_0^2|A_Q|^2\,, \tag{2.82}
$$

where we refer to M_0 as the soft mass which A_Q acquires. Similarly, $\mathcal{L}_2$ induces a holomorphic soft mass, which due to symmetry constraints arises only in the Higgs sector, with $Q^2 \to H_uH_d$. Furthermore, $\mathcal{L}_3$ induces soft trilinear or 'A-terms':

$$
\mathcal{L}_3 \supset \frac{F_S}{M}A_Q^3 \equiv A\cdot A_Q^3\,. \tag{2.83}
$$

Finally, the last term induces gaugino masses $M_{1/2}$,

$$
\mathcal{L}_4 \supset \frac{F_S}{M}\lambda^\alpha\lambda_\alpha \equiv M_{1/2}\lambda^\alpha\lambda_\alpha\,. \tag{2.84}
$$

A standard scenario, known as 'Gravity Mediation' has $M \sim M_P \sim 10^{18}$ GeV, a value which corresponds the scale at which one may expect quantum gravity to induce all allowed higher-dimension operators. Then one would need the SUSY breaking scale in the hidden sector to be $\sqrt{|F_S|} \sim 3\times 10^{10}$ GeV (which is sometimes referred to as an 'intermediate scale') to obtain

$$
M_0 \sim A \sim M_{1/2} \sim 1\,\text{TeV}\,. \tag{2.85}
$$

Of course, many new parameters are introduced in this way. In particular, there are as many A-terms as there are entries in the Yukawa coupling matrices, and the soft masses come as

3×3 matrices in generation space. If the scale of the soft terms (sometimes referred to as the SUSY breaking scale) is low - e.g. in the TeV range, then generic values for the soft terms are ruled out by flavor changing neutral currents (FCNCs) and other experimental signatures. Some symmetry-based model building is needed to make this scenario realistic.

It is crucial that no renormalizable couplings between hidden and visible sector are present. In particular, a superpotential term SQ^2 (or concretely SH_uH_d) would induce a Higgs mass $\sim \sqrt{|F_S|}$, destabilizing the hierarchy. Furthermore, we need a non-zero μ term for the Higgs, but it should not be too large, again to avoid a hierarchy destablization.

There is a very elegant solution to the problem of inducing a supersymmetric μ term of the same size as the (otherwise very similar) SUSY-breaking holomorphic mass term $\sim H_uH_d$. This is known as the **Giudice-Masiero mechanism** and is based on the higher-dimension couplings

$$\mathcal{L} \supset \left(\frac{1}{M} \overline{S} H_u H_d + \frac{1}{M^2} \overline{S} S H_u H_d \right) \Big|_{\theta^2 \overline{\theta}^2} . \tag{2.86}$$

They induce terms

$$\mathcal{L} \supset \frac{\overline{F}_S}{M} H_u H_d \Big|_{\theta^2} + \frac{|F_S|^2}{M^2} H_u H_d \tag{2.87}$$

at the (partial) component level. The first is the previously discussed μ term, but with a coefficient governed by the SUSY-breaking scale, $\mu \sim \overline{F}_S/M$, the second is the so-called B_μ term, the previously mentioned holomorphic mass term for the Higgs:

$$\mathcal{L} \supset B_\mu H_u H_d \qquad \text{with} \qquad B_\mu \sim |F_S|^2/M^2 . \tag{2.88}$$

Upon integrating out the F-terms of the Higgs superfields, the μ term also contributes to the quadratic Higgs scalar potential, which in total reads

$$\begin{aligned} V_2 &= (|\mu|^2 + m_{H_u}^2)|H_u|^2 + (|\mu|^2 + m_{H_d}^2)|H_d|^2 + B_\mu H_u H_d + \text{h.c.} \\ &= \begin{pmatrix} H_u \\ \epsilon \overline{H}_d \end{pmatrix}^\dagger \begin{pmatrix} |\mu|^2 + m_{H_u}^2 & B_\mu \\ B_\mu & |\mu|^2 + m_{H_d}^2 \end{pmatrix} \begin{pmatrix} H_u \\ \epsilon \overline{H}_d \end{pmatrix} . \end{aligned} \tag{2.89}$$

The second line makes it apparent that we are dealing simply with a 4×4 complex mass matrix, giving mass to the 4 scalars contained in $(H_u, \epsilon \overline{H}_d)^T$. Due to $SU(2)$ symmetry, this matrix has a 2×2 block structure and hence only two distinct eigenvalues. Electroweak symmetry breaking requires one of the eigenvalues to be negative.

An independent quartic Higgs interaction is not present in the SUSY Standard Model since no cubic Higgs superpotential is allowed. However, the D term of the $SU(2) \times U(1)$ SUSY gauge theory does the job:

$$V_4 = \frac{1}{8}(g_1^2 + g_2^2)\,(|H_u|^2 + |H_d|^2)^2 + \frac{1}{2} g_2^2 |H_u \overline{H}_d|^2 . \tag{2.90}$$

Assuming soft terms are close to the weak scale, the scalar potential $V_2 + V_4$ and its symmetry breaking structure has been analysed in great detail, but we will not discuss this. Suffice it to say that electroweak symmetry can be broken as required, both Higgs doublets generically develop

VEVs (the ratio being parametereized by $\tan\beta \equiv v_u/v_d$), and the Higgs mass is predicted in terms of this mixing angle and the gauge couplings. This is a great success, given in particular that all parameters of this model are now protected from power-divergences, the SUSY breaking and weak scale are naturally small, and the model is renormalizable and can, in principle, be valid all the way to the Planck scale. In addition, extrapolating the Standard Model gauge couplings to high energy scale, one finds that they meet rather precisely at the GUT scale $M_{GUT} \simeq 10^{16}$ GeV. This has been known since about 1990 and has provided a lot of credibility to the model.

However, the predicted Higgs mass is bounded by the Z-boson mass at tree level, which is clearly incompatible with observations. The correction need to bring the Higgs mass up to its observed value of 125 GeV can be provided by loops, but this requires a large stop quark mass or large trininear terms. This drives (again through loops) the Higgs VEV to higher values and partially spoils the success of the hierarchy problem resolution. In addition, the non-discovery of superpartners at the LHC has raised the lower limits for soft terms, also limiting the success of the hierarchy problem resolution. Thus, the phenomenological status of this model has deteriorated. From a modern prespective, it may be more appropriate to think of this not as of a weak scale model but rather as a model at a higher scale $m_{soft} \gg m_{ew}$.

This perspective implies that one integrates out all SUSY partners and the second Higgs at m_{soft} and is left with just the Standard Model below that scale. More precisely, this requires that the lowest eigenvalue of the Higgs mass matrix in (2.89) is smaller than the typical entries (which are $\sim m_{soft}^2$). This is a fine tuning of the order m_{ew}^2/m_{soft}^2 which one may have to accept. This fine tuning ensures that m_H^2 of the Standard Model Higgs, which sets the weak scale, is somewhat below the SUSY breaking scale. One may refer to this as a 'high-scale' or 'split' MSSM [43,44], and it is not implausible that such a model (or some variant thereof) arises in string theory (see e.g. [45,46]).

2.11 Supergravity - superspace approach

All that was said above must, of course, be consistently embedded in a generally relativistic framework and the resulting structure, known as supergravity, is equally elegant and unique, though technically much more complicated, than flat-space SUSY. We can only give a brief summary of results. Since we described flat SUSY using the superspace approach, let us start by noting that a similar (curved) superspace approach can also be used to derive supergravity [35,47]. For a brief discussion of this see also [48].

One starts, as before, with coordinates

$$z^M = (x^\mu, \theta^\tau, \overline{\theta}_{\dot{\tau}}) \tag{2.91}$$

with the above indices being 'Einstein indices', as in conventional general relativity. Then one introduces a vielbein, $E_A{}^M(z)$, i.e. a basis of vectors, labelled by the 'Lorentz indices'

$$A = (a, \alpha, \dot{\alpha})\,. \tag{2.92}$$

As in general relativity, one defines a connection, introduces constraints (such as the vanishing torsion constraint), and removes gauge redundancies. This is very cumbersome in the present case, but it eventually leads to a supergravity superspace action

$$S = \int d^8 z\, E\, \Omega(\Phi, \overline{\Phi}) + \int d^6 z\, \varphi^3\, W(\Phi) + \text{h.c.} \tag{2.93}$$

Here E is the determinant of the vielbein $E_A{}^M$. The latter contains a real vector superfield and an (auxiliary) chiral superfield

$$\begin{aligned} \mathcal{H}^\mu &= \theta\sigma^a\overline{\theta}e_a{}^\mu + i\overline{\theta}^2\theta^\alpha\psi_\alpha{}^\mu + \text{h.c.} + \theta^2\overline{\theta}^2 A^\mu && (2.94)\\ \varphi &= e^{-1}\left(1 - 2i\sigma^a\overline{\psi}^\mu + F_\varphi\theta^2\right) && (2.95) \end{aligned}$$

with $e = \det(e_a{}^\mu)$. We thus have the component fields

$$e_a{}^\mu(x)\,, \qquad \psi_\alpha{}^\mu(x)\,, \qquad A^\mu(x)\,, \qquad F_\varphi(x)\,. \tag{2.96}$$

Here the first is the familar vielbein of Einstein's theory, and the last two are auxiliaries (some authors use the notation $B(x) \equiv F_\varphi(x)$). The crucial new feature is a physical, propagating spin-(3/2) field, called the **gravitino**, which is the superparter of the vielbein (or equivalently of the metric or graviton). The z integrations are over the full or half of the Grassmann part of superspace, as in the flat case. The argument Φ stands for as many chiral superfields, containing matter degrees of freedom, as one wants. The function Ω is real.

It goes far beyond the scope of these notes to derive the component action. However, to get a glimpse of what is going on, we can consider the flat-space limit:

$$e_a{}^\mu = \delta_a{}^\mu\,, \quad \psi_\alpha{}^\mu = 0\,, \quad A^\mu = 0\,, \quad \varphi = 1 + \theta^2 F_\varphi\,. \tag{2.97}$$

Then the action takes the form

$$S = \int d^8 z\, \varphi\overline{\varphi}\, \Omega(\Phi, \overline{\Phi}) + \int d^6 z\, \varphi^3\, W(\Phi) + \text{h.c.} \tag{2.98}$$

From this, integrating out F_Φ and F_φ, one can straightforwardly obtain the supergravity scalar potential. To be specific, one finds the potential in the Brans-Dicke-frame. This is so because, in the curved case, one would have also found

$$S \supset \int d^4x\, \sqrt{g}\, \frac{1}{2} M_P^2\, \mathcal{R} \cdot \frac{\Omega(\Phi, \overline{\Phi})}{3}\,, \tag{2.99}$$

i.e., the Einstein-Hilbert term in the Brans-Dicke frame. Rescaling the metric to absorb the factor $\Omega/3$, one arrives at an Einstein-frame curvature term together with the supergravity scalar potential

$$V = e^{K/M_P^2}\left(K^{i\overline{\jmath}}(D_i W)(\overline{D}_{\overline{\jmath}}\overline{W}) - 3|W|^2/M_P^2\right) \tag{2.100}$$

where

$$D_i W = \partial_i W + K_i W \tag{2.101}$$

and

$$K = -3M_P^2 \ln[-\Omega/(3M_P^2)] \qquad \text{or} \qquad \Omega = -3M_P^2 \exp[-K/(3M_P^2)]\,. \tag{2.102}$$

This goes together with conventional kinetic terms for the fields Φ^i, based on the supergravity Kahler metric $K_{i\bar{\jmath}}$. We have given all of the above keeping M_P explicit to make it easy to see that the flat space limit, $M_P \to \infty$, takes us back to previous formulae. In particular, one can see that the first term in (2.100) correponds to the familar F-term scalar potential while second term is supergravity-specific. It is non-zero even if W is just a number and thus allows for the introduction of a cosmological constant, albeit only a negative one. This is consistent with the fact that the Poincare superalgebra can be generalized to Anti-de-Sitter but not to de-Sitter space.

In practice, one mostly works with the above formulae in units in which $M_P = 1$. This is much more economical and we will do so from now on.

Let us note that, among many other terms, one has

$$\mathcal{L} \supset -e^{K/2}\overline{W}\psi_\mu\sigma^{\mu\nu}\psi_\nu + \text{h.c.}\,, \tag{2.103}$$

which implies a gravitino mass

$$m_{3/2} = e^{K_0/2}W_0\,, \tag{2.104}$$

where W_0 and K_0 are the vacuum values of W and K. We will suppress the indices '0' from now on since it will be clear from the context whether the vacuum value or some other dynamical value is meant. Supersymmetry breaking is, as before, governed by non-zero VEVs of (some of) the F-terms,

$$F^i = e^{K/2}D^iW = e^{K/2}K^{i\bar{\jmath}}\overline{D_{\bar{\jmath}}W}\,. \tag{2.105}$$

Realistically, we have $\lambda = V_0 \simeq 0$ (the non-zero meV-scale value is negligible compared to particle-physics scales). Hence, the positive-definite F-term piece and the negative $|W|^2$ piece compensate with high precision in the formula for V. We thus have

$$|F| \sim e^{K/2}|W| \qquad \text{and hence} \qquad m_{3/2} \sim |F|\,. \tag{2.106}$$

Here $|F|$ is the length of the vector F^i, calculated using the Kahler metric $K_{i\bar{\jmath}}$. We note, however, that this is in Planck units and, reinstating M_P, one has $m_{3/2} \sim |F|/M_P$. Thus, if one takes the hidden-sector F very low, near the weak scale (as is in principle consistent with our SUSY-breaking discussion), the gravitino can still be very light. This, however, requires that it couples to Standard Model fields only very weakly.

We note that the SUSY solution to the weak-scale hierarchy problem works as before: All that we said remains valid since we are working at an EFT scale $\mu \ll M_P$ and the rigid limit (supplemented by the gravitino, if it is sufficiently light) can be used. The non-renormalization theorem extends to the W of supergravity. What is more, the presence of higher-dimension operators which was central in the communication of SUSY breaking from hidden to visible sector can be argued to be generic in the supergravity context: After all, the theory is non-renormalizable, so all in principle allowed operators are expected to be present with M_P-suppression. Also, the non-linear structure of Ω expressed in terms of K suggests such operators. The corresponding, very generic way of SUSY breaking mediation (through Planck suppressed higher-dimension operators) is called **gravity mediation**.

2.12 Supergravity - component approach

Before closing this chapter, we should note that we only discussed the superfield approach to supergravity since it fits the previous analysis of rigid supersymmetry best. It is not the most economical or widely used approach, which is instead based on the component form of SUSY multiplets and (superconformal) tensor calculus [49].

Very briefly, the story can be told as follows: In general relativity, Lorentz symmetry becomes local. Since the SUSY parameter ξ, being a spinor, transforms non-trivially under the Lorenzt group, it would be inconsistent to consider it a global object. Instead, it must be promoted to a space-time dependent quantity,

$$\xi \quad \to \xi(x)\,, \tag{2.107}$$

such that supersymmetry becomes a gauge symmetry. But now we are clearly missing a gauge field defining the connection associated with our gauge symmetry. By analogy to

$$A_\mu(x) \quad \to \quad A_\mu(x) + \partial_\mu \epsilon(x)\,, \tag{2.108}$$

one writes

$$\psi_\mu(x) \quad \to \quad \psi_\mu(x) + \partial_\mu \xi(x)\,. \tag{2.109}$$

The new field ψ_μ is a vector-spinor, also known as gravitino. We here interpret both ξ and ψ_μ as 4-component spinors, specifically Majorana spinors. The presence of the gravitino can also be motivated in a different way:

Indeed, we are clearly missing a superpartner for the graviton. As it turns out, the right object is ψ_μ. To understand this better, we take a step back, forget about superfields, and recall the SUSY algebra with its generators Q and $\overline{Q}$ (that come on top of the Poincare generators). They have spin and hence raise or lower the spin of objects on which they act. Indeed, developing the representation theory of the SUSY Poincare algebra one finds multiplets including particles with different spin or, in the massless case, helicity. We aready know the multiplets

$$(0, 1/2) \qquad \text{and} \qquad (1/2, 1) \tag{2.110}$$

corresponding to the chiral and real superfield (or the scalar and vector multiplet). Naturally, one expects and indeed finds the multiplet

$$(3/2, 2) \tag{2.111}$$

containing gravitino and graviton. For this to be consistent, one needs the gravitino to contain 2 degrees of freedom on shell, to match those of the graviton. Indeed, the general expressions for numbers of degrees of freedom of a vector spinor, initially and after taking into account gauge redundancy, constraints and the on-shell condition, are

$$d \cdot 2^{[d/2]} \qquad \to \qquad \frac{1}{2}(d-3) \cdot 2^{[d/2]}\,. \tag{2.112}$$

Here the exponent $[d/2]$ (the integer fraction of $d/2$) characterizes the dimension of a general spinor, the reduction from d to $d-3$ is associated with gauge freedom and constraints, and the prefactor $1/2$ is the usual reduction from off-shell to on-shell degrees of freedom affecting any spinor (due to the equation of motion being first order).

We record for completeness the underlying action and equation of motion (the Rarita-Schwinger equation),

$$S = -\int d^dx\,\overline{\psi}_\mu \gamma^{\mu\nu\rho}\partial_\nu\psi_\rho \qquad \text{and} \qquad \gamma^{\mu\nu\rho}\partial_\nu\psi_\rho = 0\,, \tag{2.113}$$

although we will not have time to discuss the derivation of the physical degrees of freedom from this dynamical description. Furthermore, we should note that the modern way of deriving actions in this context is the so-called **tensor calculus**. By this one means rules for multiplying (combining) multiplets to obtain new multiplets. We saw an expample of this when we noted that $\Phi_1(y,\theta)\Phi_2(y,\theta)$, with Φ_1 and Φ_2 chiral, defines a new chiral superfield. This can be formulated without superspace, just on the basis of the components. With this methods, the full action of supergravity, including supergravity coupled to chiral and vector multiplets, can be derived. More specifically, the method of choice is 'superconformal tensor calculus', which first extends the theory to a conformal supergravity, then breaks scale invariance by a VEV and removes the extra degrees of freedom by constraints. (The non-SUSY version of this would be to replace the Planck scale by a field and then recover usual gravity by giving this field a VEV.) In fact, this superconformal method is also used in the superspace approach and we saw a trace of the field whose VEV eventually breaks scaling symmetry in the chiral compensator $\varphi(y,\theta)$ of (2.98).

Let us end with part of the general 4d supergravity action (the full action being given e.g. in [35, 49]). The input are three functions, the (real) Kahler potential K, the holomorphic superpotential W and the (also holomorphic) gauge-kinetic function f_{ab}:

$$\begin{aligned}
\frac{1}{\sqrt{g}}\mathcal{L} \;=\;& \frac{1}{2}R + \epsilon^{\mu\nu\rho\sigma}\overline{\psi}_\mu\overline{\sigma}_\nu D_\rho\psi_\sigma + K_{i\bar{\jmath}}\left[(D_\mu\phi^i)(D^\mu\overline{\phi}^{\bar{\jmath}}) - i\overline{\chi}^{\bar{\jmath}}\overline{\sigma}^\mu D_\mu\chi^i\right] \\
&+(\mathrm{Re}f_{ab})\left[-\frac{1}{4}F^a{}_{\mu\nu}F^{b\,\mu\nu} - \overline{\lambda}^a\overline{\sigma}^\mu D_\mu\lambda^a\right] + \frac{1}{4}(\mathrm{Im}f_{ab})F^a{}_{\mu\nu}\tilde{F}^{b\,\mu\nu} \\
&-e^{K/2}\left[\left(\overline{W}\psi_\mu\sigma^{\mu\nu}\psi_\nu + \frac{i}{\sqrt{2}}(D_iW)\chi^i\sigma^\mu\overline{\psi}_\mu + \frac{1}{2}(D_iD_jW)\chi^i\chi^j\right) + \text{h.c.}\right] \\
&-V_F - V_D + \{\,\text{further fermionic terms}\,\}\,.
\end{aligned} \tag{2.114}$$

Here

$$\begin{aligned}
D_iW &= W_i + K_iW \\
D_iD_jW &= W_{ij} + K_{ij}W + K_iD_jW + K_jD_iW - K_iK_jW - \Gamma_{ij}{}^kD_kW\,.
\end{aligned} \tag{2.115}$$

We already know the F-term potential

$$V_F = e^K\left(K^{i\bar{\jmath}}(D_iW)(\overline{D}_{\bar{\jmath}}\overline{W}) - 3|W|^2\right). \tag{2.116}$$

The D-term potential has until now only been given implicitly and in a special case. More generally, it reads (cf. [50] for a very compact discussion)

$$V_D = \frac{1}{2}[(\mathrm{Re}\,f)^{-1}]^{ab}D_aD_b\,. \tag{2.117}$$

To define the D terms, we recall that the scalars parameterize a Kahler manifold which, to be gauged, must have some so-called (holomorphic) Killing vector fields

$$X_a = X_a^i(\phi)\frac{\partial}{\partial\phi^i}\,. \tag{2.118}$$

They define the direction in which the manifold can be mapped to itself by the gauge transormation corresponding to the index a. They also appear in the general formula for the covariant derivatives:

$$(D_\mu\phi)^i = \partial_\mu\phi^i - A_\mu^a X^i(\phi)\,. \tag{2.119}$$

Now, the D terms are defined as real solutions of the differential equations (the Killing equations)

$$X_a{}^i = -iK^{i\bar{\jmath}}\,\frac{\partial D_a(\phi,\overline{\phi})}{\partial\overline{\phi}^{\bar{\jmath}}}\,. \tag{2.120}$$

Mathematically, they are the Killing potentials. They can be given explicitly as

$$D_a = iK_i X_a{}^i + \xi_a\,, \tag{2.121}$$

where the ξ_a are so-called supergravity FI terms. The latter are only allowed for abelian generators and they correspond to gauged R-symmetries.

The terms we omitted when writing the action involve kinetic mixings between matter fermions, gauginos, and gravitino (which requires gauge symmetry or SUSY breaking) as well as 4-fermion-terms and couplings between fermions and the gauge field strength.

3 String Theory: Bosonic String

3.1 Strings – basic ideas

What we have achieved so far is not entirely satisfactory: Supersymmetry (more precisely, the broader framework of supergravity) offers a partial solution to the weak-scale hierarchy problem. Partial refers to the fact that SUSY partners have not been discovered (yet?) and hence some fine-tuning is probably needed after all. Supergravity is needed to combine this with general relativity, but it does not help with the cosmological constant problem, which unavoidably shows up in this context. Technically, the cosmological constant can be anything in supergravity: It can be negative due to the $-3|W|^2$ term, or positive due to a dominant $|DW|^2$ term (with SUSY sponteneaously broken). It is also affected by UV divergences since (in spite of the non-renormalization theorems for W), the Kahler potential K is loop corrected. Moreover, the UV problems of gravity (all operators being generated at the scale M_P – i.e. formal 'non-renormalizability') is not resolved by the prefix 'super'.

The string idea is illustrated in Fig. 4 and states simply that point-particles should be replaced by little loops of *fundamental string*. This might help with UV divergences (especially in gravity) since the interaction point is gone. Hence, when calculating a loop, there is no way in which this loop can go to zero size by the (e.g. two) interaction points becoming infinitely close. Some of the many standard textbooks are [51–55].

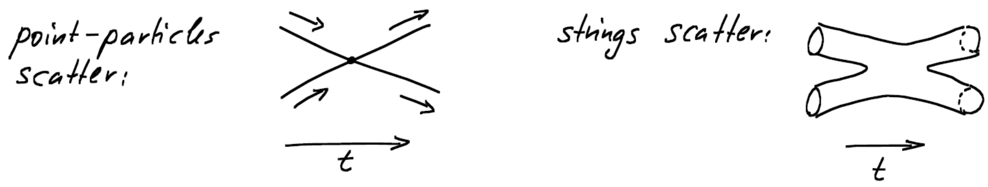

Figure 4: Point particle scattering vs. string scattering.

But before discussing scattering, we will of course have to understand how a single string loop moves through space (in other words, how its **worldsheet** is embedded in **target-space**, more precisely, in target spacetime), see Fig. 5. Before doing so, let us consider the more familiar case of a point particle, cf. Fig. 6. The embedding of the worldline γ in target space is specified by the set of functions $X^\mu(\tau)$ and the natural action is

$$S = -m \int_\gamma ds \qquad \text{with} \qquad ds^2 = -\eta_{\mu\nu} dX^\mu dX^\nu \qquad \text{and} \qquad dX^\mu = \dot{X}^\mu \, d\tau \,. \tag{3.1}$$

More explicitly, this action can be written as

$$S = -m \int d\tau \sqrt{-\eta_{\mu\nu} \dot{X}^\mu \dot{X}^\nu} \,. \tag{3.2}$$

One can easily check that this is reparameterization invariant under $\tau \to \tau' = \tau'(\tau)$ and that the non-relativistic limit is

$$S = \int dt \left(\frac{m}{2} \overline{v}^2 - m \right) \,. \tag{3.3}$$

Much more could be said about this simple and familiar system (see e.g. [56]), but for now this will suffice to motivate the Nambu-Goto action for the string.

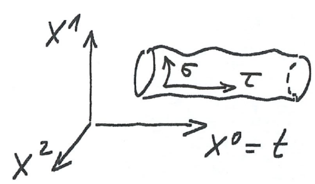

Figure 5: String moving through target space.

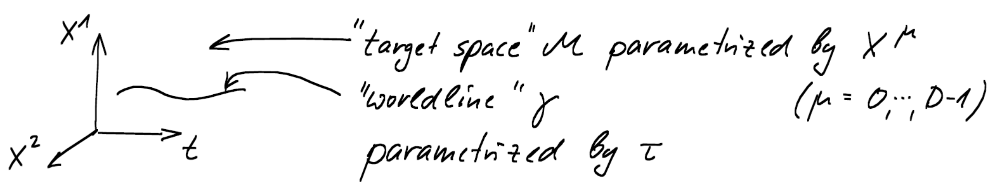

Figure 6: Point particle through target space.

In complete analogy to the point particle, the **Nambu-Goto action** for the bosonic string measures the surface area of the worldsheet embedded in target space:

$$S_{NG} = -T \int_\Sigma df\,. \tag{3.4}$$

To write this more explicitly, one parametrizes the worldsheet by (cf. Fig. 5)

$$\xi \equiv (\xi^0, \xi^1) \equiv (\tau, \sigma)\,. \tag{3.5}$$

The surface area is nothing but the volume of the 2d manifold, parameterized by ξ, measured with the induced metric G_{ab}. The latter is defined by

$$ds^2 = \eta_{\mu\nu} dX^\mu dX^\nu = \eta_{\mu\nu}\, \partial_a X^\mu\, \partial_b\, X^\nu d\xi^a\, d\xi^b \equiv G_{ab}\, d\xi^a d\xi^b\,. \tag{3.6}$$

Hence

$$S_{NG} = -T \int_\Sigma d^2\xi\, \sqrt{-G} \qquad \text{with} \qquad G \equiv \det(G_{ab})\,. \tag{3.7}$$

The prefactor T determines the tension of the string.

Due to the square root, the system is hard to quantize on the basis of this action. Instead, one uses the classically equivalent **Polyakov action**

$$S_P = -\frac{T}{2} \int_\Sigma d^2\xi\, \sqrt{-h}\, h^{ab} \partial_a X^\mu\, \partial_b X^\nu\, \eta_{\mu\nu}\,. \tag{3.8}$$

Here we introduced a new degree of freedom – the worldsheet metric h_{ab}. To see the equivalence, one integrates out h_{ab} by solving its equations of motion

$$0 = \delta_h \left[\sqrt{-h} h^{ab} G_{ab}\right] = -\frac{\delta h}{2\sqrt{-h}} h^{ab} G_{ab} + \sqrt{-h}\, \delta h^{ab}\, G_{ab}\,. \tag{3.9}$$

Next, one makes use of the fact that, for a generic matrix A, one has

$$\delta\,(\det A)/(\det A) = \delta \ln(\det A) = \delta\, \mathrm{tr}\, \ln(A) \tag{3.10}$$

and hence

$$\delta\,(\det A) = (\det A)\, \mathrm{tr}(A^{-1}\, \delta A) = -(\det A)\, \mathrm{tr}(A\, \delta A^{-1})\,. \tag{3.11}$$

Applying this to δh, the equation of motion for h_{ab} becomes

$$0 = \delta h^{ab}\, \left[\frac{h}{2\sqrt{-h}} h_{ab}\, h^{cd} G_{cd} + \sqrt{-h} G_{ab}\right] \tag{3.12}$$

or, using the identity $h/\sqrt{-h} = -\sqrt{-h}$,

$$\frac{1}{2} h_{ab}\, h^{cd} G_{cd} = G_{ab}\,. \tag{3.13}$$

It is solved by $h_{ab} = \alpha G_{ab}$ for any α. Inserting this in the Polyakov action,

$$S_P = -\frac{T}{2} \int d^2\xi\, \sqrt{-h}\, h^{cd} G_{cd} = -\frac{T}{2} \int d^2\xi\, \sqrt{-\alpha^2 G}\, 2\alpha^{-1} = S_{NG}\,, \tag{3.14}$$

one obtains the Nambu-Goto action.

At this point, jumping somewhat ahead, we can sketch what will follow: The Polyakov action describes simply a 2d field theory of D free scalars, living on a cylider ($S^1 \times$ [Time]). This is a quantum mechanical system and its states have the interpretation of particles living in the D-dimensional target spacetime. Consistency will require $D = 26$, and the spectrum will contain a massless graviton and other massless (as well as many heavy) fields. However, it will also contain a particle with negative mass squared, a tachyon. Thus, the vacuum of the 26d gravitational field theory which this bosonic string describes is unstable. This instability problem will be cured if we move on to the superstring (based on a 2d supersymmetric worldsheet theory). The target spacetime will then have to be 10d and contact with the real world will be based on compactifying this 10d supergravity to 4d. The last step means considering geometries $M_6 \times \mathbb{R}^4$, with M_6 a compact 6d manifold.

3.2 Symmetries, equations of motion, gauge choice

It is convenient to think of the worldsheet theory as of a 2d QFT with metric h_{ab} and D free scalars X^μ:

$$S_P = -\frac{T}{2}\int d^2\xi\,\sqrt{-h}\,(\partial X)^2\;, \qquad (\partial X)^2 = h^{ab}(\partial_a X^\mu)(\partial_b X^\nu)\eta_{\mu\nu}\,. \tag{3.15}$$

The three key symmetries of this theory are

(1) Diffeomorphism: $\xi^a \;\to\; \xi'^a(\xi^0,\xi^1)$.

(2) Poincare symmetry: $X^\mu \;\to\; X'^\mu = \Lambda^\mu{}_\nu X^\nu + V^\nu$ with $\Lambda \in SO(1, D-1)$.

(3) Weyl rescalings: $h_{ab}(\xi) \;\to\; h'_{ab}(\xi) = h_{ab}(\xi)\exp[2\omega(\xi)]$, with ω an arbitrary real function.

The first and second are obvious and follow immediately from the structure of our worldsheet action. It is noteworthy that target-space Poincare symmetry is an interal symmetry from the worldsheet perspective. The third is a specialty of the string. In other words, for a similar theory of moving p-branes, parameterized by $\xi^0, \xi^1, \cdots, \xi^p$, this symmetry does not exist unless $p = 1$.

To move on, it is convenient to use the energy-momentum tensor,

$$T^{MN} = \frac{2}{\sqrt{-g}}\cdot\frac{\delta S}{\delta g_{MN}} \qquad \text{or, equivalently} \qquad T_{MN} = \frac{-2}{\sqrt{-g}}\cdot\frac{\delta S}{\delta g^{MN}}\,, \tag{3.16}$$

which takes the form $T_{MN} = \mathrm{diag}(\rho, p, \cdots, p)$ for an isotropic fluid. On the string worldsheet, a slightly different normalization is common:

$$T^{ab} = \frac{-4\pi}{\sqrt{-h}}\cdot\frac{\delta S_P}{\delta h_{ab}}\,. \tag{3.17}$$

One easily checks that

$$T^{ab} = -\frac{1}{\alpha'}\left(G^{ab} - \frac{1}{2}h^{ab}(h^{cd}G_{cd})\right)\,, \tag{3.18}$$

where we also introduced the **Regge slope**

$$\alpha' \equiv \frac{1}{2\pi T}\,. \tag{3.19}$$

The latter is a different way to parameterize the string tension. It goes back to the early days of string theory, when the focus was on string theory as a model of hadronic physics. This is nicely explained in the first chapter of [51].

It follows both from our discussion in the last section as well as from the general definition of T^{ab} that the equation of motion of h_{ab} is

$$T^{ab} = 0\,. \tag{3.20}$$

Moreover, tracelessness holds as an identity, i.e. independently of whether the field configuration obeys the equations of motion:

$$T^a{}_a = 0 \qquad \text{for any} \qquad h_{ab}\,. \tag{3.21}$$

The reader should convince herself that this generally follows from symmetry (3). Finally, the equations of motion of X are

$$\Box X^\mu = 0 \qquad \text{with} \qquad \Box = D^a \partial_a\,. \tag{3.22}$$

It is crucial for what follows that **diffeomorphisms** and **Weyl rescalings** are (by definition) not just symmetries but **gauge redundancies**. This allows one to work in the **flat gauge**,

$$h_{ab} = \text{diag}(-1, 1)\,. \tag{3.23}$$

Indeed, very superficially one can argue as follows: A 2d metric contains 3 real functions. Diffeomorphisms and Weyl rescalings also contain $2 + 1 = 3$ real functions. Hence, it should be possible to bring h_{ab} to any desired form.

In somewhat more detail, one can explicitly check that

$$\sqrt{-h'}\mathcal{R}[h'] = \sqrt{-h}\,(\mathcal{R}[h] - 2D^2\omega) \qquad \text{for} \qquad h'_{ab} = e^{2\omega} h_{ab}\,. \tag{3.24}$$

Now, starting from any metric h, one may try to solve the equation $2D^2\omega = \mathcal{R}$. This can always be achieved (in non-compact space with localized source $\mathcal{R}$) since it only requires the inversion of the Klein-Gordon operator. Without prove, we simply state that this holds also on the cylinder, which is our case of interest. For more details, see e.g. [53].

Once $2D^2\omega = \mathcal{R}$ is solved, one can Weyl rescale h using the solution ω. The resulting metric will have vanishing Ricci scalar and, since in $d = 2$

$$\mathcal{R}_{abcd} = \frac{1}{2}(h_{ac}h_{bd} - h_{ad}h_{bc})\mathcal{R}\,, \tag{3.25}$$

it will be flat. More precisely, the worldsheet is a flat metric manifold and hence there exist coordinates in which the metic is manifestly flat in the sence of (3.23).

Let us now focus on a flat worldsheet and on the corresponding equations of motion

$$(\partial_\tau^2 - \partial_\sigma^2)X^\mu = 0\,. \tag{3.26}$$

It is convenient to use light-cone coordinates $\sigma^\pm = \tau \pm \sigma$, such that

$$ds^2 = -d\tau^2 + d\sigma^2 = -d\sigma^+ d\sigma^- \qquad \text{and} \qquad h_{++} = h_{--} = 0\;, \qquad h_{+-} = h_{-+} = -\frac{1}{2} \tag{3.27}$$

and

$$\Box = -4\partial_+\partial_- \qquad \text{with} \qquad \partial_\pm = \frac{\partial}{\partial\sigma^\pm}\,. \tag{3.28}$$

The equations of motion take the form

$$\partial_-\partial_+ X^\mu = 0 \tag{3.29}$$

and have the general solution

$$X^\mu(\sigma^+,\sigma^-) = X_L^\mu(\sigma^+) + X_R^\mu(\sigma^-)\,, \tag{3.30}$$

being further constrained by $X^\mu(\tau,\sigma) = X^\mu(\tau,\sigma+\pi)$, cf. Fig. 7. Here we have used the reparameterization freedom to set the circumference of the cylinder to π. This is a convention used in many string theory textbooks, in particular in [51] which we mostly follow.

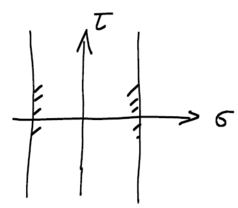

Figure 7: The strip with periodic boundary conditions on which the X^μ field theory lives.

Periodicity of X^μ implies periodicity of $\partial_+X^\mu = \partial_+X_L^\mu$ and of $\partial_-X^\mu = \partial_-X_R^\mu$. The latter depend only on σ^+ and σ^- respectively and can therefore be represented as Fourier series in these two variables:

$$\partial_+X_L^\mu \;\sim\; \text{const.}_L + \sum_{n\neq 0} f_{L,n}e^{-2in\sigma^+} \quad , \qquad \partial_-X_R^\mu \;\sim\; \text{const.}_R + \sum_{n\neq 0} f_{L,n}e^{-2in\sigma^-}\,. \tag{3.31}$$

Returning to X_L^μ and X_R^μ by integration, the exponentials remain exponentials and the constants translate into linear terms. Moreover two integration constants appear. Hence, with a choice of prefactors dictated by convention, one finds the general solution or **mode decomposition**

$$X_L^\mu \;=\; \frac{1}{2}x^\mu + \frac{l^2}{2}p^\mu\sigma^+ + \frac{il}{2}\sum_{n\neq 0}\frac{1}{n}\,\tilde{\alpha}_n^\mu\, e^{-2in\sigma^+} \tag{3.32}$$

$$X_R^\mu \;=\; \frac{1}{2}x^\mu + \frac{l^2}{2}p^\mu\sigma^- + \frac{il}{2}\sum_{n\neq 0}\frac{1}{n}\,\alpha_n^\mu\, e^{-2in\sigma^-}\,. \tag{3.33}$$

Here we introduced $l = \sqrt{2\alpha'}$, the so-called string length. One should be aware that the precise definition (the numerical prefactor) may vary from author to author and from context to context.

We note that the coefficients of the two terms linear in σ^+ and σ^- are forced to be equal by the periodicity of X^μ. Their value specificies the center-of-mass momentum of the string. By contrast, the constants $x^\mu/2$ are chosen to be equal by convention. It is only their sum that has physical meaning, characterizing the position of the center of mass at $\tau = 0$. Reality of X^μ implies that x^μ and p^μ are real, consistently with their physical meaning which we pointed out above. The oscillator modes have to satisfy

$$(\alpha_n^\mu)^* = \alpha_{-n}^\mu\,. \tag{3.34}$$

3.3 Open string

It will later on be crucial to also consider open strings. We introduce them already now since they are in fact a simpler version of the closed string – they basically carry half of the degrees of freedom. Instead of a cylinder, one now has to think of a strip (parameterized transversely by $\sigma \in (0, \pi)$) embedded in target space, cf. Fig. 8.

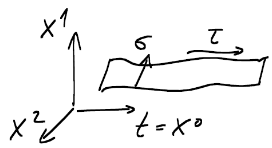

Figure 8: Open string.

The variation of the action,

$$\delta S = \frac{1}{2\pi\alpha'} \int d^2\sigma \, (\partial^2 X) \cdot \delta X - \frac{1}{2\pi\alpha'} \int d\tau \int_0^\pi d\sigma \, \partial_\sigma(\partial_\sigma X \cdot \delta X) \,, \tag{3.35}$$

now includes boundary terms. Indeed, while the first term vanishes if the equations of motion are obeyed, the second gives

$$-\frac{1}{2\pi\alpha'} \int d\tau \, (\partial_\sigma X^\mu) \cdot \delta X_\mu \Big|_{\sigma=0}^{\sigma=\pi} . \tag{3.36}$$

To avoid introducing new degrees of freedom living at the boundary, we need this to vanish. This can be achieved by two different types of boundary conditions,

$$\partial_\sigma X^\mu = 0 \quad \text{(Neumann)} \,, \qquad\qquad \delta X_\mu = 0 \quad \text{(Dirichlet)} \,. \tag{3.37}$$

In the first case the string end moves freely (no momentum is lost at the end of the string), in the second it is confined to lie in a fixed hyperplane. For example (cf. Fig. 9), one can enforce Neumann boundary conditions for X^0, X^2 and Dirichlet boundary conditions for X^1. One is then dealing with an open string living on a **D-brane** (with D referring to Dirichlet) filling out the X^0 and X^2 directions of target spacetime. Specifically, if a brane fills out p spatial dimensions, i.e. if it is a p-dimensional object in the usual, spatial sense, one calls it a **Dp-brane**. For target space to be stationary, branes always have to fill out the time or X^0 direction. This, of course, does not contribute to their dimensionality as a spatial object. However, in spacetime a Dp-brane is $(p+1)$-dimensional object.

We also note that configurations with various, also intersecting branes are permitted, cf. Fig. 10. Jumping ahead, we record that, analogously to the closed string states containing the target-space graviton, the open string states contain a massless vector particle: a $U(1)$ gauge boson. Thus, on every Dp brane one has localized $(p+1)$-dimensional gauge theory. One a stack of N D-branes, N^2 such states, since each string begin or end on any one of these N coincident branes. This gives rise to a $U(N)$ gauge theory. If branes or brane stacks intersect, then the string living at the intersection (as in the last picture in Fig. 10) gives rise to states (target-space particles or fields) which are charged under the two gauge groups corresponding

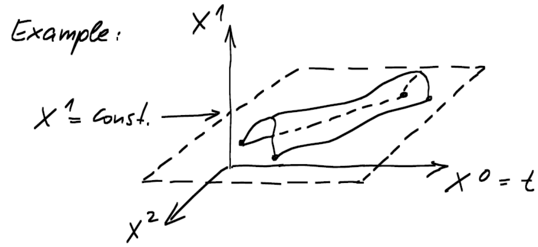

Figure 9: Open string living on a D1-brane filling out the X^2 direction.

Figure 10: Various brane configurations with strings attached.

to the two branes. This is how Standard Model matter fields arise in some of the simplest phenomenologically interesting string models – the so-called **intersecting brane models**.

What is interesting for us at the moment is that the mode decomposition of the open string is simpler than that of the closed string. Indeed, while one needs sines and cosines (or equivalently exponentials) to Fourier decompose a periodic function, on a interval one can do with just sines or just cosines. Technically, one may say (and it is easy to demonstrate this explicitly) that, for the open string, the left and right-moving modes are identified. Explicitly, for the case of Neumann boundary conditions, one has

$$X^\mu = x^\mu + l^2 p^\mu \tau + il \sum_{n\neq 0} \frac{1}{n} \alpha_n^\mu e^{-in\tau} \cos(n\sigma)\,. \tag{3.38}$$

Thus, it is often simpler to discuss the open string and then 'double' the result to go over to the closed case.

We also note that the case of Neumann boundary conditions for *all* X^μ should actually be viewed as a situation with spacetime filling branes. Thus, open strings generally end on D-branes.

3.4 Quantization

We will only present the old covariant approach, briefly commenting on light-cone and modern covariant approach (path integral or BRST quantization) at the end. The starting point is the flat-gauge Polyakov action wich, breaking 2d covariance, can be written as

$$S = \frac{1}{4\pi\alpha'} \int d^2\sigma (\dot{X}^2 - X'^2) \qquad \text{with} \qquad d^2\sigma = d\tau d\sigma\,. \tag{3.39}$$

Here we have left the index μ and its contraction implicit. Nevertheless, the above describes D free bosons and we have to keep in mind that one of them (X^0) has a wrong-sign kinetic term.

The canonical variables are

$$X^\mu \qquad \text{and} \qquad \Pi_\mu = \frac{\partial \mathcal{L}}{\partial \dot{X}^\mu} = \frac{1}{2\pi\alpha'}\,\dot{X}_\mu\,, \tag{3.40}$$

with equal-time commutation relations

$$[\hat{\Pi}_\mu(\tau,\sigma),\,\hat{X}^\nu(\tau,\sigma')] = -i\delta(\sigma-\sigma')\delta_\mu{}^\nu\ , \qquad [\hat{X}^\mu,\hat{X}^\nu] = [\hat{\Pi}_\mu,\hat{\Pi}_\nu] = 0\,. \tag{3.41}$$

Promoting our previous mode decomposition of X^μ (and a corresponding decomposition of Π^μ) to the operator level, one finds

$$[\hat{p}^\mu,\hat{x}^\nu] = -i\eta^{\mu\nu}\ , \qquad [\hat{\alpha}^\mu_m,\hat{\alpha}^\nu_n] = m\,\delta_{m+n}\eta^{\mu\nu}\ , \qquad [\hat{\tilde{\alpha}}^\mu_m,\hat{\tilde{\alpha}}^\nu_n] = m\,\delta_{m+n}\eta^{\mu\nu}\,, \tag{3.42}$$

where

$$\delta_{m+n} \equiv \delta_{m+n\,,0}\,. \tag{3.43}$$

We will drop the hats from now on, assuming that it will always be clear from the context whether the operator or the classical variable is meant.

As usual in quantum mechanics, we now need a Hilbert space representation of our operator algebra. Given the non-trivial commutation relations of p and x, we can only choose one of them to be diagonal. Since we are interested in a particle interpretation of string states, it is natural to choose p and write

$$\mathcal{H} = \sum_p \mathcal{H}(p)\,, \tag{3.44}$$

where $\mathcal{H}(p)$ is the eigenspace of the operators $\{\hat{p}^0,\cdots,\hat{p}^{D-1}\}$ with eigenvalues $\{p^0,\cdots,p^{D-1}\}$.

We now focus on the subspace corresponding to one particular value of p and rewrite the mode-algebra acting on it:

$$[\alpha^\mu_m,\alpha^\nu_n] = m\,\delta_{m+n}\eta^{\mu\nu} \qquad\qquad \to \qquad\qquad [\alpha^\mu_m,\alpha^{\nu\,\dagger}_n] = |m|\delta_{m,n}\eta^{\mu\nu}\,. \tag{3.45}$$

We see that we are dealing simply with a very large set of oscillators, labelled by μ and $m>0$. We define a vacuum state and find the Fock space:

$$\mathcal{H}(p) = \text{Span}\Big\{\,\alpha^\mu_m\alpha^\nu_n\cdots|0,p\rangle\ \Big|\ \text{any number of }\alpha\text{s; any }\mu,\nu,\cdots;\text{ any }m,n,\cdots>0\Big\}\,. \tag{3.46}$$

The situation we arrive at is very similar to the initial step of Gupta-Bleuler quantization of electrodynamics: There, on account of the vector index of A_μ and the non-positive-definite metric $\eta_{\mu\nu}$, the Fock space includes negative norm states. They are removed by a physical state condition or constraint, related to the gauge invariance of the theory. Here, the same issue arises due to the vector index of α^μ_m. As will become clear momentarily, the resolution is also similar.

We fixed part of the gauge freedom by eliminating h_{ab}. The corresponding equation of motion was $T_{ab}=0$, which now has to be implemented as a constraint. It is convenient to do this in light-cone coordinates, convincing oneself first that

$$T^a_a = 0 \qquad \Leftrightarrow \qquad T_{+-} = 0\,. \tag{3.47}$$

Since the trace vanishes identically, one only needs to enforce the constraints

$$T_{++} = T_{--} = 0\,. \tag{3.48}$$

It is straightforward to check that

$$T_{++} = (\partial_+ X_L)\cdot(\partial_+ X_L) \qquad \text{and} \qquad T_{--} = (\partial_- X_R)\cdot(\partial_- X_R) \tag{3.49}$$

and that, using the mode decomposition, the Fourier modes of the these quantities read

$$L_m \equiv \frac{1}{4\pi\alpha'}\int_0^\pi d\sigma\, T_{--}\, e^{-2im\sigma} = \frac{1}{2}\sum_{n=-\infty}^{\infty} \alpha_{m-n}\cdot\alpha_n \tag{3.50}$$

$$\tilde{L}_m \equiv \frac{1}{4\pi\alpha'}\int_0^\pi d\sigma\, T_{++}\, e^{2im\sigma} = \frac{1}{2}\sum_{n=-\infty}^{\infty} \tilde{\alpha}_{m-n}\cdot\tilde{\alpha}_n\,. \tag{3.51}$$

Here we also used the simplifying notation

$$\alpha_0^\mu = \tilde{\alpha}_0^\mu = \frac{l}{2}p^\mu\,. \tag{3.52}$$

For the open string, one defines

$$\tilde{L}_m \equiv \frac{1}{2\pi\alpha'}\int_0^\pi d\sigma\, T_{++}\, e^{2im\sigma} = \frac{1}{2}\sum_{n=-\infty}^{\infty} \tilde{\alpha}_{m-n}\cdot\tilde{\alpha}_n\,, \tag{3.53}$$

with T_{--} being a dependent quantity. One can also check that

$$H = L_0 + \tilde{L}_0 \quad \text{(closed string)}\ , \qquad\qquad H = L_0 \quad \text{(open string)}\,. \tag{3.54}$$

Note that $\alpha_0^\mu = lp^\mu$ for the open string.

One can check that the operators L_m (with or without tilde) satisfy the **Virasoro algebra**

$$[L_m, L_n] = (m-n)L_{m+n} + A(m)\delta_{m+n}\ , \qquad \text{with} \qquad A(m) = (m^3 - m)D/12\,. \tag{3.55}$$

Here the term proportional to D is called the anomaly term and D is the central charge. Note that this term depends on a possible additive redefinition of L_0, which is related to the ordering ambiguity present in all the terms of type $\alpha_{-k}\alpha_k$ in L_0. The form given above assumes normal ordering, i.e. $\langle 0,0|L_0|0,0\rangle = 0$.

The classical part of this algebra is called **Witt algebra** and is satisfied by the differential operators

$$D_m = ie^{im\theta}\frac{d}{d\theta}\,, \tag{3.56}$$

which generate diffeomorphisms on an S^1 parameterized by $\theta \in (0, 2\pi)$. These remarks are the beginning of a long and important chapter of a proper string theory course – 2d conformal field theory. However, we are not going to discuss this, such that a few comments will have to suffice:

When we fixed the gauge (diffeomorphisms and Weyl scalings), a residual gauge freedom was left. It consists of diffeomorphisms under which the metric changes only by Weyl scaling.

Now it is useful to insist on the point of view that, after going the flat gauge, we are in a fixed-background QFT and coordinate reparameterizations are forbidden. From this perspective, the residual gauge freedom noted above corresponds to space-time dependent translations of the field configuration which preserve angles, i.e., conformal transformations (Fig. 11). Our theory is invariant under those and hence is a conformal field theory or CFT [57–62]. The Virasoro algebra is the corresponding symmetry algebra. It is clear that conformal transformations can be generated as spacetime dependent translations. Given that T_{ab} generates translations, we are not surprised to find that the Fourier modes of T_{ab} are the desired symmetry generators. It is also natural that the Witt algebra, as introduced above, is the classical counterpart.

Figure 11: Illustration of a conformal mapping of a given field configuration to a new one.

The conformal symmetry just introduced is a central tool in developing string theory and, in particular, in deriving scattering amplitudes, loop corrections etc. We will have no time for this. But it may be useful to note that, when studying CFTs in their own right, the anomaly term or, equivalently, a non-zero central charge do not represent a problem. However, in string theory the conformal symmetry is part of an underlying gauge symmetry and this term must vanish. It indeed does, in the so-called critical dimensions, but to see this one needs to do the gauge fixing more carefully, introducing Fadeev-Popov ghosts. They cancel the central charge coming from the scalars.

Returning to our main line of development, we now want to be more explicit about the physical state condition. As in QED, it is sufficient to demand that the 'annihilator part' of the constraint vanishes on physical states, i.e. $L_m|\,phys\,\rangle = 0$ for $m \geq 0$. But it turns out that, at this point, a divergence present in the definition of L_0 has to be resolved. This has to do with operator ordering.

Indeed, our definition so far was

$$(L_0)_{tot} = \frac{1}{2}\sum_{n=-\infty}^{\infty} \alpha_{-n}\alpha_n = \frac{1}{2}\alpha_0^2 + \frac{1}{2}\sum_{n\neq 0}\alpha_{-n}\alpha_n\,. \tag{3.57}$$

We gave this quantity an index for 'total' since we are going to separate the normal ordered part from it in a moment. We also note that the ordering of the creation and annihilation operators used above comes directly from the original definition

$$(L_0)_{tot} = H_{tot} = \frac{1}{4\pi\alpha'}\int_0^\pi d\sigma\,(\dot{X}^2 + X'^2)\,. \tag{3.58}$$

Here, for simplicity, we think of the open string or, equivalently, just the right-moving part of the closed string.

To evaluate a constraint like $(L_0)_{tot}|\, phys \,\rangle = 0$, we want to work instead with a normal-ordered operator. Hence, we define

$$L_0 \equiv \frac{1}{2}\sum_{n=-\infty}^{\infty} :\alpha_{-n}\alpha_n: = \frac{1}{2}\alpha_0^2 + \sum_{n>0}\alpha_{-n}\alpha_n \,. \tag{3.59}$$

Note that this **supercedes** our previous definition of (3.50). The two definitions differ by a divergent normal ordering constant,

$$(L_0)_{tot} = L_0 - a \qquad \text{with} \qquad a = -\frac{1}{2}(D-2)\sum_{n=1}^{\infty} n \,, \tag{3.60}$$

following simply from $(1/2)(\alpha_n\alpha_n^\dagger + \alpha_n^\dagger\alpha_n) = \alpha_n^\dagger\alpha_n + n/2$. The prefactor $(D-2)$ counts the number of oscillators that contribute. The direct calculation gives, of course, D, but we have corrected this to $(D-2)$ on account of the wrong-sign scalar X^0. This is necessary since this wrong sign-scalar is associated with negative norm states, which are connected with the still unfixed (residual) gauge freedom. The latter corresponds to conformal transformations or reparameterizations of type $\sigma_+ \to \sigma'_+ = \sigma'_+(\sigma)$ and $\sigma_- \to \sigma'_- = \sigma'_-(\sigma)$, which preserve the flat gauge. One can fix this further gauge freedom (light-cone gauge and light-cone quantization), which manifestly gets rid of all oscillators except the $(D-2)$ transversal ones. Alternatively, one can use the Fadeev-Popov method and introduce ghosts, which will precisely cancel the two modes which we removed by hand.

We have no time to discuss this in detail, but the reader familiar with QED will immediately see the analogy with the photon case: Of the formally four degrees of freedom associated with the vector A_μ, only two transverse modes contribute to physical quantities like Casimir effect or vacuum free energy. This happens for exactly the same reason as here and to see it explicitly in a covariant QED calculation one also needs ghosts.

The simplest way to move on is to use ζ function regularization:

$$\left\{\sum_{n=1}^{\infty} n\right\}_{reg.} = \lim_{s\to -1}\left(\sum_{n=1}^{\infty} n^{-s}\right) = \lim_{s\to -1}\zeta(s) = \zeta(-1) = -\frac{1}{12}\,. \tag{3.61}$$

This is of course quite formal and not very satisfying. Since the result is important, we want to spend some time to explain why the normal-ordering constant does in fact have a physical and a-priori finite definition. To see this, we give the infinite strip on which our 2d field theory lives a proper, physical width: $\pi \to \pi R$. Then we have

$$H_{tot} = \frac{1}{2R}\sum_{n=-\infty}^{\infty}\alpha_{-n}\alpha_n + \pi R\lambda \,. \tag{3.62}$$

Crucially, we have here also introduced a cosmological constant counterterm.

We are now dealing with a standard QFT problem – the calculation of the total energy of a 2d theory on a strip of width πR. The sum over zero modes has a UV divergence, to be regularized by a introducing a cutoff scale Λ. A cutoff dependence must also be assigned to the counterterm, $\lambda \to \lambda(\Lambda)$. Its form is determined by the requirement that the divergence for

$\Lambda \to \infty$ cancels. Moreover, no finite ambiguity will arise since we know from Weyl invariance that the renormalized cosmological constant must vanish. Thus, we have

$$H_{tot} = \frac{1}{R}\left(\frac{1}{2}\alpha_0^2 + \sum_{n>0}\alpha_{-n}\alpha_n\right) \; + \; \lim_{\Lambda\to\infty}\left[\frac{D-2}{2R}\left\{\sum_{n-1}^{\infty} n\right\}_{\Lambda} + \pi R \lambda(\Lambda)\right]. \tag{3.63}$$

A very intuitve way of regularizing this is to think in terms of physical modes with momenta $k_n = n/R$ and to multiply the contribution of each mode by $\exp(-k_n/\Lambda)$. It is then a straightforward exercise to do the summation, find the appropriate counterterm $\lambda(\Lambda)$, and to obtain the finite result

$$H_{tot} = \frac{1}{R}\left(\frac{1}{2}\alpha_0^2 + \sum_{n>0}\alpha_{-n}\alpha_n\right) \; - \; \frac{D-2}{24R}\,. \tag{3.64}$$

The physical interpetation of this finite correction is clear: This is a one-loop **Casimir energy**, associated with the finite size of the space on which the QFT lives.[9]

Returning to our stringy convention with $R = 1$ and to the notation L_0 instead of H, we have

$$(L_0)_{tot} = L_0 - a \qquad \text{with} \qquad L_0 = \frac{1}{2}\alpha_0^2 + \sum_{n>0}\alpha_{-n}\alpha_n \qquad \text{and} \qquad a = \frac{D-2}{24}\,. \tag{3.65}$$

The physical state condition hence reads ($m \geq 0$)

$$(L_m - a\delta_m)|\, phys\,\rangle = 0\,, \tag{3.66}$$

where it is crucial to remember that L_0 is, by definition, normal ordered.

3.5 Explicit construction of physical states – open string

We start with the open-string worldsheet **vacuum**,

$$|0,p\rangle\;, \qquad\qquad \text{defined by} \qquad\qquad \hat{p}^\mu|0,p\rangle = p^\mu|0,p\rangle\,. \tag{3.67}$$

For $m > 0$, our physical state condition reads

$$L_m|0,p\rangle = \frac{1}{2}\sum_n \alpha_{m-n}\alpha_n|0,p\rangle = 0\,, \tag{3.68}$$

which is satisfied for any p since in any term of this sum either $n > 0$ or $m - n > 0$. Thus, there is always an annihilator involved, giving zero if applied to the vacuum.

By contrast, the $m = 0$ condition is non-trivial, giving

$$\left(\frac{p^2}{2} + \sum_{n>0}\alpha_{-n}\alpha_n - a\right)|0,p\rangle = 0\,, \tag{3.69}$$

[9] Demonstrating that this result is in fact independent of the precise form of the regularizing function, $\exp(-k/\Lambda) \to f(k/\Lambda)$ is not entirely trivial. See e.g. [5] for a discussion of the corresponding 4d problem.

where we have set $l = 1$ such that $\alpha_0 = p$. With $M^2 = -p^2$, this translates into

$$M^2 = -2a\,. \tag{3.70}$$

Thus, not any p is allowed. Rather, the above mass-shell condition must be satisfied.

Moving to the **first excited level**, we have to consider states

$$\zeta_\mu \alpha^\mu_{-1}|0,p\rangle\,, \tag{3.71}$$

with a polarizaton vector ζ. The mass shell condition now reads

$$0 = (L_0 - a)\zeta_\mu \alpha^\mu_{-1}|0,p\rangle = \left(\frac{p^2}{2} + \alpha_{-1}\cdot\alpha_1 - a\right)\zeta_\mu\alpha^\mu_{-1}|0,p\rangle = \left(\frac{p^2}{2} + 1 - a\right)\zeta_\mu\alpha^\mu_{-1}|0,p\rangle\,, \tag{3.72}$$

implying

$$M^2 = 2(1-a)\,. \tag{3.73}$$

Of the L_m conditions with $m > 0$, now the first also becomes non-trivial:

$$0 = L_1\,\zeta\cdot\alpha_{-1}|0,p\rangle = \left(\frac{1}{2}\sum_n \alpha_{1-n}\cdot\alpha_n\right)\,\zeta\cdot\alpha_{-1}|0,p\rangle\,. \tag{3.74}$$

Of the various terms in the sum, only those can contribute where $n \leq 1$ and $1-n \leq 1$. This occurs only for $n = 0, 1$, such that we find

$$0 = \frac{1}{2}(\alpha_1\cdot\alpha_0 + \alpha_0\cdot\alpha_1)\,\zeta\cdot\alpha_{-1}|0,p\rangle = \zeta\cdot\alpha_0|0,p\rangle = \zeta\cdot p|0,p\rangle\,. \tag{3.75}$$

The implication is that the polarization has to be transverse. We also need the norm of the state, which is

$$\langle 0,p|(\zeta_\mu\alpha^\mu_{-1})^\dagger(\zeta_\nu\alpha^\nu_{-1})|0,p\rangle = \langle 0,p|0,p\rangle\zeta_\mu\zeta^\mu = \zeta^2\,. \tag{3.76}$$

At the so-called **second excited level**, one has to analyse states of the form

$$(\epsilon_{\mu\nu}\alpha^\mu_{-1}\alpha^\nu_1 + \epsilon_\mu\alpha^\mu_{-2})|0,p\rangle\,, \tag{3.77}$$

but we will not do so.

Instead, we summarize the results by focussing on the **first excited level**:

(A) For $a > 1$ we have $M^2 < 0$. This means that p is spacelike and hence timelike ζ with $\zeta\cdot p = 0$ exist. Thus, there are allowed states with $\zeta^2 < 0$ and hence negative norm. This is excluded.

The deeper is reason for the problem is that a Weyl anomaly arises, which can only be cured by considering a background which is not simply flat D-dimensional Minkowski space. This is know as **Supercritical string theory** and may (in the supersymmetric case) nevertheless be relevant phenomenologically, although this is not established.

(B) For $a = 1$, we have $M^2 = -p^2 = 0$, implying that the $(D-1)$ independent ζs which satisfy $\zeta\cdot p = 0$ fall into two classes:

One longitudinal, $\zeta||p$, with zero norm.

$(D-2)$ transverse, which are spacelike and give rise to positive-norm states. One may think e.g. of $p=(1,1,0,\cdots,0)$ and $\zeta_i=(0,0,\cdots,0,1,0,\cdots,0)$, with unity at position $1+i$ and $i\in\{1,2,\cdots,(D-2)\}$. This is consistent with Gupta-Bleuler quantization QED. It gives the correct description of a gauge theory with a massless vector. This case is known as **Critical string theory** and we will completely focus on this case, with the **critical dimension** $D=26$, in what follows.

(C) For $a<1$ we have $M^2>0$ for the first excited (and all higher) levels. Thus except possibly for the vacuum state, this case is practically not very interesting. It is not inconsistent at the present level of analysis (giving rise to a massive vector with $(D-1)$ positive norm states). Problems, possibly solvable, arise in the interacting theory. This is known as the **Subcritical string**. The Weyl anomaly is also present, as in the supercritical case. Together, cases (A) and (B) are known as non-critical string theory.

We close by mentioning that the overall picture in the critical case is just like in gauge theory quantization: We have restricted our Fock space by imposing the physical state condition. The resulting space has no negative norm states, but so-called null states are still present. The actual positive-definite **Hilbert** space is constructed as a quotient

$$\mathcal{H}_0\equiv\mathcal{H}_{phys}/\mathcal{H}_{null}\,. \tag{3.78}$$

The mass-shell condition, originating from $(L_0-1)|\,phys\,\rangle=0$, can be written as

$$M^2=-p^2=2(N-1)\qquad\text{with}\qquad N\equiv\sum_{n>0}\alpha_{-n}\alpha_n\,. \tag{3.79}$$

The operator N or its expectation value is called the **level**. We have found a **tachyon** at level 0, a massless vector at level 1, and we could have found massive string excitations at level 2 and higher. The tachyon corresponds to the statement that our assumed 26d Minkowski vaccum is unstable since a scalar with negative mass squared is present. It will decay by tachyon condensation, which is an interesting subject of research. But we will not discuss this since we use the bosonic string only as a toy model get ready for the superstring.

3.6 Explicit construction of physical states – closed string

A repetition of the analysis of the previous section will again single out the case $a=1$ or $D=26$. We focus right away on this case, recalling however that the number of operators and constraints is now doubled. We rewrite

$$(L_0-a)|\,phys\,\rangle=0\qquad,\qquad(\tilde L_0-a)|\,phys\,\rangle=0 \tag{3.80}$$

as

$$(L_0-\tilde L_0)|\,phys\,\rangle=0\qquad,\qquad(\tilde L_0+\tilde L_0-2a)|\,phys\,\rangle=0\,. \tag{3.81}$$

We recall that

$$L_0=\frac{\alpha_0^2}{2}+N=\frac{p^2}{8}+N\,, \tag{3.82}$$

since for $l = 1$ one finds $\alpha_0 = p/2$ in the closed-string case. Analogous equations hold for the left-movers. With this, the physical state conditions become

$$(N - \tilde{N})|\,phys\,\rangle \qquad \text{and} \qquad (p^2/4 + N + \tilde{N} - 2)|\,phys\,\rangle\,, \tag{3.83}$$

known as **level matching** and **mass shell conditions** respectively. The latter is also frequently given as

$$M^2 = 4(N + \tilde{N} - 2)\,. \tag{3.84}$$

Now one proceeds systematically, level by level, as before. At the **vacuum** level one again finds a tachyon,

$$|0, p\rangle\ , \qquad M^2 = -8\,. \tag{3.85}$$

At the first excited level, due to the level matching condition, both α_{-1} and $\tilde{\alpha}_{-1}$ have to be used:

$$\xi_{\mu\nu}\alpha^{\mu}_{-1}\tilde{\alpha}^{\nu}_{-1}|0, p\rangle\ , \qquad M^2 = 0\,. \tag{3.86}$$

Note that, as before, one really has $M^2 = 4(1 + 1 - 2a)$, such that masslessness follows only for $a = 1$, i.e. in the critical dimension. At the first excited level, the L_1 and $\tilde{L}_1$ constraints are non-trivial. They read

$$\xi_{\mu\nu}p^{\mu} = 0 \qquad \text{and} \qquad \xi_{\mu\nu}p^{\nu} = 0\,. \tag{3.87}$$

It is also easy to check that the norm of our states is

$$\langle\,phys\,|\,phys\,\rangle \sim \xi_{\mu\nu}\xi^{\mu\nu}\,, \tag{3.88}$$

which is always non-negative if the physical state conditions are satisfied.

To classify the states, it is helpful to think of the polarization tensor literally as of an element in the tensor product of two copies of $\mathbb{R}^D$,

$$\xi^{\mu\nu} = \sum_{ab} v^{\mu}_{(a)} \otimes v^{\nu}_{(b)}\,. \tag{3.89}$$

In analogy to the standard treatment of the photon, one chooses a basis $v_{(a)}$ with one element $v_{(0)} \sim p$, one lightlike element $v_{(1)}$ with non-zero product with p, and $D - 2$ spacelike elements. Of these, only the spacelike and $v_{(0)}$ are allowed to appear, hence we have $(D-1)^2$ physical basis states. Furthermore, $(D - 2)^2$ of them (those built from spacelike vectors only) have positive norm. The rest corresponds to gauge freedom.

Choosing $p \sim v_{(0)} \sim (1, 1, 0, \cdots, 0)$ and $v_{(1)} \sim (1, -1, 0, \cdots, 0)$, we see explicitly how products of the $(D - 2)^2$ transverse vectors form a basis for the transverse polarizations ξ_t, which correspond to the lower-right corner of the matrix ξ:

$$\xi = \left(\begin{array}{ccccc} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \cdot & \cdot & & \xi_t & \\ 0 & 0 & & & \end{array}\right)\,. \tag{3.90}$$

The transverse physical polarizations ξ_t transform under $SO(D-2)$, the group of rotations in the spacelike hyperplane transverse to p. This is called 'little group' – the subgroup of $SO(1, D-1)$ leaving p invariant. This rank-2 tensor representation is not irreducible but decomposes into symmetric, antisymmetric and trace part, corresponding to 3 different fields of the D-dimensional field theory which the string describes from the target space perspective. They are:

(1) The graviton $G_{\mu\nu}$, with $(D-1)(D-2)/2-1$ d.o.f.s (note that for $D=4$ this correctly reproduces the known result of 2 d.o.f.s).

(2) The Kalb-Ramon field or antisymmetric tensor $B_{\mu\nu}$, with $(D-2)(D-3)/2$ d.o.f.s.

(3) The dilaton ϕ, with 1 d.o.f.

We could go on to discuss excited states, but all we will need to know is that there they form a tower with increasing mass and the number of states at each consecutive level grows extremely fast. The mass spacing is $\Delta M^2 \sim 1/\alpha'$.

3.7 The 26d action

We are only interested in the critical case, $D = 26$, and we focus on the closed string (for more details see e.g. [52]). It is immediate to write down a quadratic-level action for the above fields (to be supplemented by the tachyon which, as we know, has negative mass squared and make the 26d Minkowski-space solution unstable). Assuming that one also knows how to compute scattering amplitudes, one can supplement this action by interaction vertices and write down the full, non-linear expression at the 2-derivative level. It reads (suppressing the tachyon):

$$S = \frac{1}{\kappa^2}\int d^{26}x\sqrt{-G}e^{-2\phi}\left[\mathcal{R}[G] - \frac{1}{12}H_{\mu\nu\rho}H^{\mu\nu\rho} + 4(\partial\phi)^2\right]. \tag{3.91}$$

where

$$H = dB\,, \tag{3.92}$$

in complete analogy with $F = dA$ in the 1-form case.

Many important comments have to be made. First, it is apparent that the value of κ^2 can be changed by a shift of ϕ. Thus, we can for example define $\kappa^2 = c\alpha'^{12}$, with some numerical constant c. Then the choice of the background value of ϕ determines the 26d Planck mass relative to the mass of the excited string modes, which is $\sim 1/\sqrt{\alpha'}$. It also governs the perturbativity of the theory, i.e. the importance of string loops, as we will discuss further down.

Second, the apparently wrong sign of the dilaton-kinetic term is misleading. Indeed, the above is called the string-frame action (similar to what is called the Brans-Dicke frame in the non-string context) and one can go to the Einstein frame by the Weyl rescaling

$$G_{\mu\nu} = \tilde{G}_{\mu\nu}\, e^{-\phi/6}\,. \tag{3.93}$$

The result is

$$S = \frac{1}{\kappa^2}\int d^{26}x\sqrt{-\tilde{G}}\left[\mathcal{R}[\tilde{G}] - \frac{1}{12}e^{-\phi/3}\, H_{\mu\nu\rho}H^{\mu\nu\rho} - \frac{1}{6}(\partial\phi)^2\right]. \tag{3.94}$$

In this frame, the Planck mass is manifestly fixed and the mass of the excited states changes with varying dilaton background.

Third, this is the first (but not the last) time we encounter a higher-form gauge theory. So it may be useful to remind the reader of some basics (a standard summary of conventions can be found in the Appendix of Volume II of [52]). It is convenient to think not just of an antisymmetric field $A_{\mu_1 \cdots \mu_p}$ but of a differential form,

$$A_p = \frac{1}{p!} A_{\mu_1 \cdots \mu_p} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_p} \,. \tag{3.95}$$

Our present case $p = 2$ with $B_{\mu_1 \mu_2} = A_{\mu_1 \mu_2}$ is part of the more general structure of such gauge theories.

One should think of the dx^μ as basis vectors of the dual tangent space (the cotangent space) of a manifold, such that

$$dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta^\mu_\nu \,. \tag{3.96}$$

Higher p-forms take their values in the p-fold exterior product (the antisymmetric part of the tensor product) of the cotanget space. This is symbolized by the wegde, e.g.

$$dx^1 \wedge dx^2 = dx^1 \otimes dx^2 - dx^2 \otimes dx^1 \,. \tag{3.97}$$

It generalizes to

$$dx^1 \wedge \cdots \wedge dx^p = p! \, dx^{[1} \otimes \cdots \otimes dx^{p]} \,, \tag{3.98}$$

where $[\cdots]$ stands for antisymmetrization. The implication is, for example, that

$$A_p(\partial_1, \cdots, \partial_p) = A_{[1 \cdots p]} = A_{1 \cdots p} \,. \tag{3.99}$$

Consistent with the above, one formally defines the product of two forms

$$(A_p \wedge B_q)_{\mu_1 \cdots \mu_{p+q}} = \frac{(p+q)!}{p!q!} A_{[\mu_1 \cdots \mu_p} B_{\mu_{p+1} \cdots \mu_{p+q}]} \,. \tag{3.100}$$

Crucially, the natural map from functions (0-forms) to 1-forms,

$$d : \, f \;\mapsto\; df = \partial_\mu f \, dx^\mu \qquad \text{with} \qquad df(\partial_\mu) \equiv \partial_\mu f \,, \tag{3.101}$$

has a generalization to higher forms:

$$(dA_p)_{\mu_1 \cdots \mu_{p+1}} = (p+1) \partial_{[\mu_1} A_{\mu_2 \cdots \mu_{p+1}]} \,. \tag{3.102}$$

By its very definition, a p-form provides, at every point, a totally anstisymmetric map from the p-fold tensor product of the tangent space to the real numbers. Thus, it can be used to define the volume of an infinitesimal parallelepiped (with orientation, i.e. ordering of the vectors by which it is spanned) at any point of the manifold. This gives rise to the possibility of integrating a p-form over a finite p-dimensional submanifold:

$$V(C_p) = \int_{C_p} A_p \,. \tag{3.103}$$

With this, interpreting A_p as a physical gauge potential, one has the invariant field strength and the gauge transformation

$$F_{p+1} = dA_p \qquad \text{and} \qquad A_p \;\to\; A_p + d\chi_{p-1}\,. \tag{3.104}$$

The natural lagrangian is $|F_{p+1}|^2 \sim F_{\mu_1\cdots\mu_{p+1}}F^{\mu_1\cdots\mu_{p+1}}$ and the natural coupling to charged objects is

$$S_{matter} \sim \int_{\Sigma_p} A_p\,. \tag{3.105}$$

This is completely analogous to electrodynamics, where the matter coupling is the integral of A_1 along the worldline of the electron. Here, it is the integral of A_p along the p-dimensional worldvolume Σ_p of a $(p-1)$-brane. (Recall the convention that the variable p in the term Dp-brane counts only the spatial dimensions.)

In our case at hand, a charged objects suitable as a source for B_2 is already present in the theory we have so far developed: It is the fundamental string itself. Thus, the term

$$\int_{\Sigma_2} B_2 \tag{3.106}$$

has to be added both to our 10d action and to our worldsheet action for the string. If B_2 is non-zero, this changes our 2d theory and its quantization.

Similarly, we see that (3.91), with $G_{\mu\nu} = \eta_{\mu\nu}$, $\phi = 0$ and $B = 0$ describes the solution in the background of which our fundamental string propagates. This is the 2d theory one is easily able to quantize. But clearly other solutions for this 10d action exist and the string can be quantized in their background as well. The 2d theory is then much more complicated, e.g. through

$$\int d^2\sigma\sqrt{-h}h^{ab}(\partial_a X^\mu)(\partial_b X^\nu)\eta_{\mu\nu} \qquad \to \qquad \int d^2\sigma\sqrt{-h}h^{ab}(\partial_a X^\mu)(\partial_b X^\nu)G_{\mu\nu}(X)\,. \tag{3.107}$$

We see that this theory now ceases to be free or quadratic in the fields. For example, if near $X = 0$ we can write

$$G_{\mu\nu} = \eta_{\mu\nu} + c\cdot(X^1)^2\eta_{\mu\nu} + \cdots\,, \tag{3.108}$$

we encouter a quartic interaction vertex in the worldsheet theory. Similarly, a non-zero $B_2 = B_2(X)$ adds new terms to the worldsheet lagrangian. In particular, the X dependence of B_2 leads to new interaction terms in the theory of scalars X^μ living on the worldsheet.

Before closing this section, we should discuss the role of the dilaton ϕ from the worldsheet perspective. This field is related to the Einstein-Hilbert term,

$$\int d^2\sigma\sqrt{-h}\,\mathcal{R}\,, \tag{3.109}$$

of the worldsheet action. At first sight, this term is clearly allowed. It respects all symmetries of the worldsheet, including in particular Weyl invariance. A more careful analysis reveals, however, that it can be written as a total derivative and hence does not affect the equations of motion. Indeed, following the standard derivation of Einstein's equations from the Einstein-Hilbert action, one finds

$$\delta_h \int d^2\sigma\sqrt{-h}\,\mathcal{R} = \int d^2\sigma\,\left(\mathcal{R}_{ab} - \frac{1}{2}h_{ab}\mathcal{R}\right)\,\delta h^{ab} \;+\; \text{boundary terms}\,. \tag{3.110}$$

But in $d = 2$ one has

$$\mathcal{R}_{ab} - \frac{1}{2} h_{ab} \mathcal{R} = 0 \tag{3.111}$$

as an identity. This follows from the symmetries of the Riemann tensor which, as already noted above, can be expressed in terms of the Ricci scalar. Thus, the worldsheet Einstein-Hilbert term does not change under continuous deformations of the worldsheet metric. Its integral can however be non-zero, measuring topological features of the worldsheet (see below).

Now, comparing the dynamics described by the target-space action given above and the role of the Einstein-Hilbert term on the worldsheet (we will see more details of this further down), one can establish that ϕ has to be identified with the coefficient of the worldsheet Einstein-Hilbert term:

$$S_P \supset \frac{1}{4\pi} \int d^2\sigma \sqrt{-h}\, \mathcal{R}\, \phi \,. \tag{3.112}$$

As before, if $\phi = \phi(X)$ is non-constant, new interactions are introduced into the worldsheet theory.

As a final remark, we note that backrounds solving the 26d equations of motion are precisely those in which the Weyl invariance on the propagating string worldsheet remains unbroken. In this sence, the 2d theory can be used to directly determine the 26d dynamics, without calculating scattering amplitudes and comparing them to 26d EFT vertices.

4 String Theory: Interactions and Superstring

Before we can see what the string-theoretic UV completion of gravity has to say about the real world, a few more formal developments are necessary. First, we want to understand at least in principle how scattering amplitudes and loop effects are calculated. Second, we need to introduce fermions and get rid of the tachyon.

4.1 State-operator correspondence

Before discussing scattering amplitudes and loops, a few more words about the worldsheet theory after gauge fixing are necessary. We learned that this is a CFT and we will here work with the Euclidean version of this theory. The symmetries of the CFT include angle-preserving deformations of the worldsheet. For example, we can map our fundamental cylinder corresponding to the propagation of the string to the z-plane,

$$z = r\, e^{i\varphi} \in \mathbb{C} \,, \tag{4.1}$$

such that time runs radially and circles of constant r correspond to constant-time cuts through our cylinder (cf. Fig. 12). The reader is invited to consider the explicit map $z = \exp(iw)$ and identify a strip in the w-plane (with periodic boundary conditions, i.e. a cylinder) that is mapped to the z-plane in the desired way.

Next, let us recall that a state in a 4d QFT may, analogously to the Schrödinger wave function of quantum mechanics, be described by a **Schrödinger wave functional**,

$$\Psi :\, \phi \,\mapsto\, \Psi[\phi, t] \in \mathbb{C} \,. \tag{4.2}$$

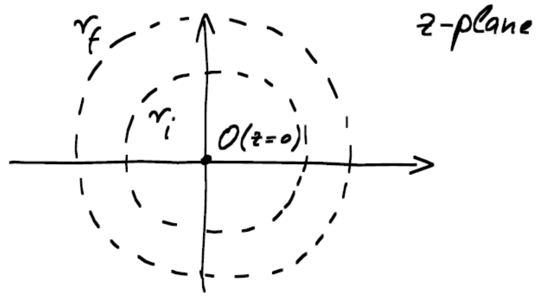

Figure 12: String propagation mapped to the z-plane. The part of the cylinder between initial time τ_i and final time τ_f corresponds to the annulus (ring) between r_i and r_f.

Here $\phi : \overline{x} \mapsto \phi(t, \overline{x}) \in \mathbb{R}$ is a field configuration at fixed time. The evolution of such states is described by the QFT version of the Feynman path integral.

In our context, a string state at time τ_i is then represented, in the radial representation, by a wave functional

$$\Psi_i[X_i, r_i]\,. \tag{4.3}$$

Here X_i stands for any of the possible field configurations $X_i^\mu(r_i, \varphi)$. The wave functional obtained by Hamiltonian evolution at radial time r_f reads

$$\Psi_f[X_f, r_f] = \int DX_i \int_{X_i}^{X_f} DX\, e^{-S[X]}\, \Psi_i[X_i, r_i]\,. \tag{4.4}$$

Here the labels X_i and X_f of the integral mean that we integrate over field configurations $X(r, \varphi)$ satisfying $X(r_i, \varphi) = X_i(\varphi)$ and $X(r_f, \varphi) = X_f(\varphi)$.

Now, consider the limit in which our evolution starts at $\tau_i = -\infty$, corresponding to $r_i = 0$ or $z = 0$. In this limit, we can write (4.4) as

$$\Psi_f[X_f, r_f] = \int^{X_f} DX\, e^{-S[X]} \lim_{r_i \to 0} \Psi_i[X(r_i, \varphi), r_i] = \int^{X_f} DX\, e^{-S[X]}\, O(z = 0)\,. \tag{4.5}$$

Here, in first step, we have absorbed the integral over X_i in the integral over X, dropping the initial boundary condition. In the second step, we have introduced the operator O in the CFT, i.e., some expression involving $X(0)$ and its derivatives.

By the above procedure, we have understood how a given state, in our case the state defined by Ψ_i, specifies an operator. The opposite direction is obvious: Clearly, Equation (4.5) may be interpreted as describing the evolution of some state, defined by O, from $\tau = -\infty$ to τ_f.

Thus, we now know how to associate a CFT state with an operator and vice versa.

4.2 Scattering amplitudes

After the discussion of the previous section, it should be at least intuitively clear that the integral over fields on a cylinder can be replace by the integral over fields on the sphere, with appropriate operators inserted at the points which are mapped to $\tau = \pm\infty$. This is illustrated in Fig. 13 together with an analogous map corresponding to the 2-to-2 scattering of string states.

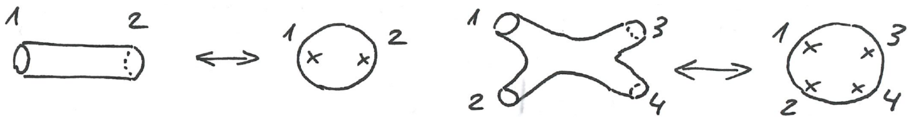

Figure 13: Identification of string propagation or scattering with compact worldsheets (in this case spheres) with operator insertions.

This leads very naturally to the following fundamental formula for n-point scattering amplitudes in string theory, which one may view as the definition of the theory:

$$\mathcal{A}_n = \sum_{g=0}^{\infty} \int \frac{Dh\, DX}{Vol_{Diff.\times Weyl}}\, e^{-S[X,h]} \int d^2z_1 \cdots d^2z_n\; V_1(z_1,\overline{z}_1) \cdots V_n(z_n,\overline{z}_n)\,. \tag{4.6}$$

Here the sum is over all compact oriented 2d manifolds (Riemann surfaces), as illustrated in Fig. 14. The terms are labelled by the genus g of the worldsheet. The integration is not only over field scalar field configurations X but also over metrics h. This definition is more fundamental and the gauge-fixed integral just over X (corresponding to the CFT language) must be carefully derived from it. The reason is that there is a non-trivial interplay between the topology of the manifold, the position of the vertex operators and the residual gauge freedom. In this process, one also has to divide out the infinite factor coming from gauge redundancies. This factor becomes manifest when one uses the Fadeev-Popov method to treat the functional integration.

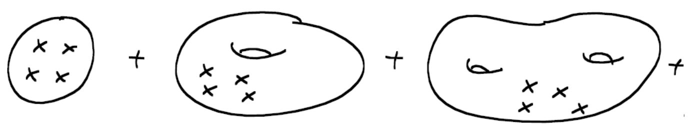

Figure 14: Contributions of worldsheets of genus zero, one and two to the four-point scattering amplitude.

The action to be used in the above is

$$S[X,h] = \frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \sqrt{-h}(\partial X)^2 + \frac{\phi}{4\pi} \int_\Sigma d^2\sigma \sqrt{-h}\mathcal{R}\,, \tag{4.7}$$

where we suppress boundary terms relevant in the open-string case. We have also assumed that the target space is flat and the dilaton constant. Restricting our attention to the oriented string, one has

$$\frac{1}{4\pi} \int_\Sigma d^2\sigma \sqrt{-h}\mathcal{R} = \chi(\Sigma) = 2 - 2g\,, \tag{4.8}$$

where χ is known as the Euler number. Thus, the second term on the r.h. side of (4.7) just supplies a factor

$$g_s^{-\chi} = g_s^{-2+2g} \qquad \text{with} \qquad g_s \equiv e^{\phi}\,. \tag{4.9}$$

The quantity g_s is known as the **string coupling**.

Finally, the so-called **vertex operators** V_i have to be chosen appropriately to reflect the physical states in the scattering of which one is interested. They can be derived from our understanding of the physical states of the quantized string and the state-operator mapping. Here, we only provide as an example the vertex operator for the tachyon of momentum k, which is basically the simplest operator that has the desired transformation properties under translations:

$$V(k,z,\overline{z}) = g_s \; : e^{ik^\mu X_\mu(z,\overline{z})} : \; . \tag{4.10}$$

The normalization by g_s ensures that that free propagation is proportional to g_s^0, tree-level 4-point-scattering to g_s^2, one-loop 4-point-scattering to g_s^4, and so on.

From here, it would be relatively straightforward to calculate some of the simplest amplitudes and loop diagrams (a.k.a. higher genus contributions) and convince oneself of the promised very soft UV behaviour and loop finiteness. But we have to move on.

4.3 World sheet supersymmetry

We need to find a string theory which desribes target space fermions and which has no tachyon. Both can be achieved by supersymmetrizing the worldsheet. As in 4d, we simply add fermionic worldsheet coordinates,

$$\sigma^a \qquad \text{`+'} \qquad \theta_\alpha \, . \tag{4.11}$$

The 2d Lorentz transformations are completely analogous to the familiar 4d case,

$$\sigma^a \;\to\; \Lambda^a{}_b \sigma^b \;, \qquad \theta_\alpha \;\to\; S_\alpha{}^\beta \theta_\beta \, , \tag{4.12}$$

with

$$\Lambda = \exp(i\epsilon^{ab} J_{ab}) \qquad \text{and} \qquad S = \exp(i\epsilon^{ab}\, \{ i[\gamma_a, \gamma_b]/4 \}) \, . \tag{4.13}$$

One may easily check that

$$\gamma^0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \;, \quad \gamma^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \tag{4.14}$$

fulfil

$$\{\gamma^a, \gamma^b\} = -2\eta^{ab} \, . \tag{4.15}$$

Of course, both $i[\gamma_a, \gamma_b]/4$ and J_{ab}, of which there is in each case only one independent element, have the same, trivial commutation relations: After all, $SO(1,1)$ is a one-parameter group. Nevertheless, the two representations are different, as the reader may want to check explicitly.

Furthermore, since S is real, one may obviously demand that the spinor is real:

$$\theta^* = \begin{pmatrix} \theta_- \\ \theta_+ \end{pmatrix}^* = \begin{pmatrix} \theta_- \\ \theta_+ \end{pmatrix} = \theta \, . \tag{4.16}$$

This is a particularly simple version of the familiar Majorana condition

$$\psi = \psi^c \equiv C\, \overline{\psi}^T \, . \tag{4.17}$$

They reader may want to consult the appendix of Volume 2 of [52] for a systematic discussion of spinors in various dimensions.

Following very closely the familiar 4d procedure, we promote our scalars to (general) superfields,

$$X^\mu \;\to\; Y^\mu(\sigma,\theta) \tag{4.18}$$

with

$$Y^\mu(\sigma,\theta) = X^\mu(\sigma) + \overline{\theta}\psi^\mu(\sigma) + \frac{1}{2}\overline{\theta}\theta\, B^\mu(\sigma)\,. \tag{4.19}$$

Here $\overline{\theta} = \theta^\dagger\gamma^0$, as in 4d. We define SUSY generators

$$Q_\alpha = \frac{\partial}{\partial\overline{\theta}^\alpha} + i(\gamma^a\theta)_\alpha\,\partial_a\,, \tag{4.20}$$

which are also Majorana spinors, and observe that they satisfy the SUSY algbera relation

$$\{Q_\alpha, \overline{Q}^\beta\} = -2i(\gamma^a)_\alpha{}^\beta\partial_a\,. \tag{4.21}$$

The SUSY transformation can be defined by

$$\delta_\xi Y(\sigma,\theta) = (\overline{\xi}Q)\,Y(\sigma,\theta)\,, \tag{4.22}$$

leading to

$$\begin{aligned} \delta_\xi X^\mu &= \overline{\xi}\psi^\mu && (4.23)\\ \delta_\xi\psi^\mu &= -i(\gamma^a\xi)\partial_a X^\mu + B^\mu\xi && (4.24)\\ \delta_\xi B^\mu &= -i\overline{\xi}\gamma^a\partial_a\psi^\mu\,. && (4.25) \end{aligned}$$

To write down SUSY-invariant actions, it is sufficient to integrate any expression in the Y^μ over the full superspace. In our case this is simply a $d^2\sigma\, d^2\theta$ integral. However, we are looking for a specific action which would serve as a generalization of the (so far flat-space) bosonic Polyakov action. For this purpose, it is convenient to introduce the supercovariant derivative

$$D_\alpha = \frac{\partial}{\partial\overline{\theta}^\alpha} - i(\gamma^a\theta)_\alpha\partial_a\,. \tag{4.26}$$

The SUSY version of our bosonic action (with $l = 1$) can then be given as

$$S = \frac{i}{4\pi}\int d^2\sigma\, d^2\theta\, (\overline{D}^\alpha Y^\mu)(D_\alpha Y_\mu) = -\frac{1}{2\pi}\int d^2\sigma\,(\partial_a X^\mu\,\partial^a X_\mu - i\overline{\psi}^\mu\partial\!\!\!/\psi_\mu - B^\mu B_\mu)\,. \tag{4.27}$$

The auxiliary field vanishes on-shell such that, in summary, we have simply added a free fermion ψ^μ for every scalar.

4.4 World sheet supergravity

The next step is to introduce gravity (more precisely, to promote the metric to a field, since gravity in the sense of a dynamical theory does not really exist $d = 2$). This implies making SUSY local, as explained earlier.

To explain this at the technical level, we first note that (since our theory contains spinors), we have to introduce a vielbein

$$h_{ab} = (e^m)_a (e^n)_b \, \eta_{mn} \,, \tag{4.28}$$

where $a, b, \cdots$ are 'curved' or 'Einstein indices' as before and $m, n, \cdots$ are 'frame' or 'Lorentz indices'. Furthermore, since the Lorentz symmetry transforming the Lorentz indices is local, we require a spin connection, to define covariant derivatives of objects with frame indices:

$$\nabla_a v^m = (\partial_a + \omega_a) v^m \qquad \text{with} \qquad \omega_a \in \; Lie(SO(1, d-1)) \,, \tag{4.29}$$

in our case with $d = 2$. It is defined by demanding covariant constancy of the vielbein,

$$0 = \nabla_a e^m{}_b = \partial_a e^m{}_b + (\omega_a)^m{}_n e^n{}_b - \Gamma^c_{ab} e^m{}_c \,, \tag{4.30}$$

where Γ stands for the usual Christoffel symbols. Clearly, the object on which ∇ acts can transform in any representation of $SO(1, d-1)$, in which case ω_a has to be taken in that representation.

With these preliminary remarks our action becomes, in the first step,

$$S_2 = -\frac{1}{2\pi} \int d^2\sigma \, e \left\{ h^{ab} (\partial_a X^\mu)(\partial_b X_\mu) - i \overline{\psi}^\mu \gamma^a \nabla_a \psi_\mu \right\} \,. \tag{4.31}$$

The index 2 stands for 'quadratic order'. We want to make this invariant under a local version of the SUSY transformations above, i.e. with $\xi \to \xi(\sigma)$. In addition, we need to define SUSY transformations of our new field, the metric or, more appropriately, the vielbein. Working at leading order in perturbations around flat space, $e^m{}_a = \delta^m{}_a$, one postulates

$$\delta_\xi e^m{}_a = -2i\overline{\xi}\gamma^m \chi_a \,. \tag{4.32}$$

Here χ_a is the gravitino, the appearance of which on the r.h. side is natural since we introduced it earlier as the superpartner of the metric. The rest of this relation is fixed (up to normalization) by covariance.

The action of (4.31) is not invariant but, since it was invariant under the global version, its non-invariance is controlled by the derivative of ξ. Thus, we have

$$\delta_\xi S_2 = \frac{2}{\pi} \int d^2\sigma \, \sqrt{-h} \, (\nabla_a \overline{\xi}) \, J^a \,, \tag{4.33}$$

where J^a is by definition the Noether current corresponding to the global version of the symmetry. Explicitly, one finds (at quadratic order in the fields)

$$J^a = \frac{1}{2} \gamma^b \gamma^a \psi^\mu \partial_b X_\mu \,, \tag{4.34}$$

known as the **supercurrent**. This can be compensated by adding a term

$$S_3 = -\frac{2}{\pi}\int d^2\sigma\,\sqrt{-h}\,\overline{\chi}_a\,J^a = -\frac{1}{\pi}\int d^2\sigma\,\sqrt{-h}\,\overline{\chi}_a\gamma^b\gamma^a\psi^\mu\partial_b X_\mu\,, \tag{4.35}$$

and introducing the transformation law

$$\delta_\xi\chi_a = \nabla_a\xi\,. \tag{4.36}$$

For obvious reasons, this method of constructing supergravity actions is known as the Noether method.

It is not yet complete: Only after a modification of $\delta_\xi\psi$ by a term proportional to the gravitino and the addition of a quartic term,

$$S_4 = -\frac{1}{4\pi}\int d^2\sigma\,\sqrt{-h}\,(\overline{\psi}\psi)\,(\overline{\chi}_a\gamma^b\gamma^a\chi_b)\,, \tag{4.37}$$

does the theory become invariant under local SUSY. We recall that the Einstein-Hilbert term is a total derivative. This matches the fact that the gravitino kinetic term is identically zero in $d=2$ (since $\gamma^{[a}\gamma^b\gamma^{c]}=0$).

Finally, the theory is still Weyl invariant, with transformation laws

$$\delta_\omega X = 0\ ,\qquad \delta_\omega e^m{}_a = \omega e^m{}_a\ ,\qquad \delta_\omega\psi = \frac{1}{2}\omega\psi\ ,\qquad \delta_\omega\chi_a = \frac{1}{2}\omega\chi_a\,. \tag{4.38}$$

Due to SUSY, this symmetry now has a fermionic counterpart, parameterized by the infinitesimal Majorana spinor η:

$$\delta_\eta X = \delta_\eta e = \delta_\eta\psi = 0\ ,\qquad \delta_\eta\chi_a = i\gamma_a\eta\,. \tag{4.39}$$

This makes our theory **super-Weyl-invariant** and, after gauge fixing, **superconformal**.

4.5 Quantization of the superstring

The large gauge symmetry (Diffeomorphisms, local Lorentz symmetry and local SUSY, super-Weyl-invariance) allow us to go to flat gauge, with trivial vielbein and vanishing gravitino. Thus, we haved to quantize the simple action

$$S = -\frac{1}{2\pi}\int d^2\sigma\,\left[(\partial_a X^\mu)(\partial^a X_\mu) - i\overline{\psi}^\mu\gamma^a\partial_a\psi_\mu\right]\,. \tag{4.40}$$

As before, the equations of motion of the fields that have been eliminated by gauge fixing must be imposed as constraints. These are

$$T_{ab} = 0\,, \tag{4.41}$$

as before, where we now have

$$T_{ab} = (\partial_a X)\cdot(\partial_b X) + \frac{i}{2}\overline{\psi}\cdot\gamma_{\{a}\partial_{b\}}\psi - \frac{1}{2}h_{ab}\left((\partial X)^2 + \frac{i}{2}\overline{\psi}\cdot\partial\!\!\!/\psi\right)\,. \tag{4.42}$$

Here the curly brackets stand for symmetrization. In addition, we can use local SUSY and super-Weyl invariance (see [53] for details) to set the gravitino χ^a to zero. But, as we have just seen in the last section, its equations of motion correspond, at leading order in χ, to the vanishing of the supercurrent:

$$(J_a)_\alpha = 0\,. \tag{4.43}$$

This is the second, new constraint.

The mode decomposition for the bosonic part is as before. To discuss the mode decomposition of the fermionic part, write

$$S = S_B + S_F \qquad \text{with} \qquad S_F = \frac{i}{\pi}\int d^2\sigma\,(\psi_- \cdot \partial_+\psi_- + \psi_+ \cdot \partial_-\psi_+)\ , \quad \text{where} \qquad \gamma^\pm = \gamma^0 \pm \gamma^1\,. \tag{4.44}$$

This explains our indexing convention in

$$\psi = \begin{pmatrix} \psi_- \\ \psi_+ \end{pmatrix} \tag{4.45}$$

since we now see that ψ_- and ψ_+ are left and right movers respectively. Due to the fermionic nature of ψ, a sign is not detectable (observables are alway built from bilinears), such that two different types of boundary conditions (known as **Ramond** and **Neveu-Schwarz**) are possible. This leads to 4 sectors:

$$\begin{array}{lcll} \psi_+(\sigma+\pi) = +\psi_+(\sigma) & ; & \psi_-(\sigma+\pi) = +\psi_-(\sigma) & \text{R-R} \\ \psi_+(\sigma+\pi) = +\psi_+(\sigma) & ; & \psi_-(\sigma+\pi) = -\psi_-(\sigma) & \text{R-NS} \\ \psi_+(\sigma+\pi) = -\psi_+(\sigma) & ; & \psi_-(\sigma+\pi) = +\psi_-(\sigma) & \text{NS-R} \\ \psi_+(\sigma+\pi) = -\psi_+(\sigma) & ; & \psi_-(\sigma+\pi) = -\psi_-(\sigma) & \text{NS-NS}\,. \end{array} \tag{4.46}$$

Note that we could not have used an arbitrary phase $\exp(i\alpha)$ instead of the sign since our spinors are real. The mode decomposition in the R-NS sector reads

$$\psi_+^\mu = \sum_{r\in\mathbb{Z}} \tilde{\psi}_r^\mu\, e^{-2ir(\tau+\sigma)} \quad , \quad \psi_-^\mu = \sum_{r\in\mathbb{Z}+\frac{1}{2}} \psi_r^\mu\, e^{-2ir(\tau-\sigma)}\,, \tag{4.47}$$

and analogously for the other 3 sectors. The reality constraint translates to the usual relation between modes with opposite frequency, e.g. $(\psi_r^\mu)^* = \psi_{-r}^\mu$.

On the open string, one only has R-R and NS-NS sectors. This is clear since one can think of the open string as coming from the closed (to be viewed as a theory on S^1) by 'modding out' a $\mathbb{Z}_2$ symmetry. In other words, one goes from S^1 to $S^1/\mathbb{Z}_2$. The $\mathbb{Z}_2$ acts by $\sigma \to -\sigma$ on the space, which means that it exchanges left and right movers in terms of fields. This would be inconsistent in a R-NS or NS-R sector. Of course, the above can be translated into the alternative picture of the superstring on an interval, with two consistent sets of boundary conditions at the two boundaries $\sigma = 0$ and $\sigma = \pi$. We leave that to the reader.

Skipping the standard steps of canonical quanitization, we immediately display the commutation relations of the oscillator modes, promoted to operators:

$$\begin{array}{rcl} [\alpha_m^\mu, \alpha_n^\nu] & = & m\,\delta_{m+n}\,\eta^{\mu\nu} \\ \{\psi_r^\mu, \psi_s^\nu\} & = & \delta_{r+s}\,\eta^{\mu\nu} \quad \text{with} \quad \left\{ \begin{array}{ll} r,s\in\mathbb{Z} & \text{(R)} \\ r,s\in\mathbb{Z}+\frac{1}{2} & \text{(NS)} \end{array}\right. \,. \end{array} \tag{4.48}$$

The different normalizations (manifest in the prefactor m and the missing prefactor r) is conventional. As before, the operators responsible for the constraints are expanded in Fourier modes,

$$L_m = \frac{1}{\pi}\int_{-\pi}^{\pi} d\sigma\, e^{im\sigma}\, T_{++} \qquad , \qquad G_r = \frac{\sqrt{2}}{\pi}\int_{-\pi}^{\pi} d\sigma\, e^{ir\sigma}\, J_+\,, \tag{4.49}$$

with

$$L_m = \frac{1}{2} : \left\{ \sum_{n\in\mathbb{Z}} \alpha_{-n}\cdot\alpha_{m+n} + \sum_{r\in\mathbb{Z}+\nu}\left(r+\frac{m}{2}\right)\psi_{-r}\cdot\psi_{m+r}\right\} : \tag{4.50}$$

$$G_r = \sum_{n\in\mathbb{Z}} \alpha_{-n}\cdot\psi_{r+n} \qquad \text{where} \qquad \nu \equiv \left\{ \begin{array}{ll} 0 & \text{(R)} \\ 1/2 & \text{(NS)} \end{array}\right. \tag{4.51}$$

These operators generate the **super-Virasoro algebra**, of which we display only the general structure:

$$[L,L] \sim L + \mathit{anomaly}\;, \qquad [L,G]\sim G\;, \qquad \{G,G\}\sim L + \mathit{anomaly}\,. \tag{4.52}$$

As before, only the annihilator-part of the classical constraints is imposed quantum-mechanically:

$$(L_m - a\delta_m)\,|phys\rangle = 0 \quad (m\geq 0)\;, \qquad\qquad G_r\,|phys\rangle = 0 \quad (r\geq 0)\,, \tag{4.53}$$

where we note that there is no normal ordering ambiguity and hence no normal ordering constant associated with G_0.

We do not repeat the derivation but simply quote the result for the normal ordering constant:

$$a = (D-2)\left(\frac{1}{24}-\frac{1}{24}\right) = 0 \quad \text{(R)} \qquad , \qquad a = (D-2)\left(\frac{1}{24}+\frac{1}{48}\right) = \frac{D-2}{16} \quad \text{(NS)}\,. \tag{4.54}$$

We see that, in the Ramond-case, the fermions precisely cancel the effect of the bosons. In the Neuveu-Schwarz case, this supersymmetric cancellation is upset by the non-trivial boundary conditions imposed on the fermions but not on the bosons.

Let us now turn concretely to the Fock space of the open-string **NS sector**: We have

$$\text{Vacuum: } |0,k\rangle \quad , \qquad \text{Creation operators: } \alpha^\mu_{-m}\;;\;\psi^\mu_{-r} \quad (m,r>0)\,. \tag{4.55}$$

The mass shell condition reads

$$0 = (L_0 - a)\,|0,k\rangle = (\alpha' p^2 + N^\alpha + N^\psi - a)\,|0,k\rangle\,, \tag{4.56}$$

where

$$N^\alpha = \sum_{m=1,2,\cdots}\alpha_{-m}\alpha_m \qquad , \qquad N^\psi = \sum_{m=\frac{1}{2},\frac{3}{2},\cdots} r\,\psi_{-r}\psi_r\,. \tag{4.57}$$

This implies that there is a scalar at level zero,

$$\alpha' M^2 = -a\,, \tag{4.58}$$

and a (target-space!) vector corresponding to the physical $\psi_{-1/2}$ excitations at level 1/2:

$$\epsilon_\mu \psi^\mu_{-1/2}|0,k\rangle \qquad \text{with} \qquad \alpha' M^2 = \frac{1}{2} - a\,. \tag{4.59}$$

In analogy to the logic of the bosonic case, we expect that $D = 10$ (with $a = 1/2$) is the critical dimension, corresponding to the vector being massless (and the scalar a tachyon, as in the bosonic string).

Next, we turn to the open-string **R-sector**, which superficially differs only very little in that

$$N^\psi = \sum_{r=0,1,2,\cdots} r\,\psi_{-r}\cdot\psi_r = \sum_{r=1,2,\cdots} r\,\psi_{-r}\cdot\psi_r\,. \tag{4.60}$$

But this number operator leads to the very peculiar situation that the ψ_0^μ do not appear in L_0 and hence do not affect the energy (mass squared) of a state. They do, however, satisfy the non-trivial algebra (D-dimensional Clifford algebra)

$$\{\psi_0^\mu, \psi_0^\nu\} = \eta^{\mu\nu}\,. \tag{4.61}$$

Hence, every mass eigenspace must carry a representation of this algebra, i.e. it must be a target-space spinor:

$$\text{Vacuum:} \quad |\alpha,k\rangle \qquad \text{with} \qquad \alpha = 1,2,3,\cdots,2^{D/2} = 32\,. \tag{4.62}$$

Since $a = 0$, this spinor is massless. To derive the critical dimension we would need to either consider heavier, excited states or involve ghosts and vanishing central charge argument. We do not do this here and only assert that the critical dimension is still $D = 10$.

4.6 GSO or Gliozzi-Scherk-Olive projection

Before constructing the 10d superstring theories which may be relevant for the real world, we need a further technical ingredient. The underlying idea is that one may always use a projection operator (an operator P with $P^2 = P$) commuting with H to reduce the Hilbert space in a consistent manner. A familiar example is the projection on symmetric and antisymmetric subspaces of $\mathcal{H}\otimes\mathcal{H}$ to define bosons and fermions in 2-particle quantum mechanics. Another example (from this course) is the projection of functions on S^1 to functions on $S^1/\mathbb{Z}^2$, which corresponds to the projection to even and odd functions and hence to the projection from closed to open string (with Dirichlet or Neumann boundary conditions). The new Hilbert space after projection is, by definition, Image(P).

Here, we focus on the open superstring and consider

$$P = \frac{1}{2}(1 + (-1)^F)\,, \qquad \text{where} \qquad F \,\equiv\, \text{Fermion number}\,. \tag{4.63}$$

This amounts to keeping only states with even F (note that F is only defined mod 2). Concretely, one defines

$$\begin{aligned} (-1)^F|0,k\rangle &= -|0,k\rangle \quad \text{(NS)} \\ (-1)^F|\alpha,k\rangle &= |\beta,k\rangle\,\Gamma_\beta{}^\alpha \quad \text{(R)} \qquad \text{where } \Gamma \equiv \Gamma^{11} \equiv \Gamma^0\Gamma^1\cdots\Gamma^9\,. \end{aligned} \tag{4.64}$$

together with

$$(-1)^F X^\mu = X^\mu (-1)^F \qquad \text{and} \qquad (-1)^F \psi^\mu = -\psi^\mu (-1)^F \,. \tag{4.65}$$

Here the appearance of $\Gamma \equiv \Gamma^{11}$ (the 10d version of γ^5) is enforced by the necessity to have $(-1)^F$ anticommute with all ψ_r^μs, including ψ_0^μ (the latter being identified with the Γ^μs).

After this projection, the tachyon is gone and the 32-component Majorana spinor has turned into a 16-component Majorana-Weyl spinor (in 10d both conditions may be imposed together). On shell and in terms of the appropriate representations of the little group SO(8), one has

$$\text{massless vector } \; (8_v) \qquad + \qquad \text{Majorana-Weyl fermion } \; (8)\,. \tag{4.66}$$

Here the symbol 8_v stands for the (defining) vector representation of $SO(8)$, the symbol 8 for the chiral Majorana spinor. We will later on also need the opposite-chirality Majorana spinor, which corresponds to an inequivalent representation. It is denoted by $8'$.

The $8_v + 8$ found above fit a 10d supersymmetric gauge theory. But we will not develop this construction since it anyway must be coupled to a closed string sector. Our purpose was only to explain the idea of this particular projection on even (or similarly on odd) fermion number states.

4.7 Consistent type II superstring theories

We now turn to the closed string case. The name 'type II' refers to the presence of two supersymmetries (equivelently two gravitini) in 10d, as will become clear in a moment. We first recall the relevant mass-shell and level-matching conditions

$$(L_0 + \tilde{L}_0)\,|phys\rangle = 0 \qquad \text{and} \qquad (L_0 - \tilde{L}_0)\,|phys\rangle = 0 \tag{4.67}$$

with

$$L_0 = \frac{\alpha'}{4}p^2 + N - \nu \;, \qquad \tilde{L}_0 = \frac{\alpha'}{4}p^2 + \tilde{N} - \tilde{\nu} \qquad \text{and} \qquad \nu/\tilde{\nu} = \left\{ \begin{array}{cc} 0 & \text{(R)} \\ 1/2 & \text{(NS)} \end{array} \right. . \tag{4.68}$$

Note that the spacing between the different mass levels differs by a factor of 4 compared to the open string. The lowest levels in the four possible sectors are

Sector	$SO(8)$ **rep.**	**mass**
NS $-$	1	tachyon
NS $+$	8_v	massless
R $-$	$8'$	massless
R $+$	8	massless ,

where $\pm$ refers to the eigenvalue of $(-1)^F$ on which one can potentially project and $8/8'$ refer to the two inequivalent spinor representations of $SO(8)$. (Of course the '1' appearing in the row of the tachyon is only intended to say that this is a scalar with a single degree of freedom – it is strictly speaking not appropriate to classify it using the little group of massless particles in 10d.) As a side remark, the existence of these in total three 8-dimensional, inequivalent representations of $SO(8)$ is related to the $\mathbb{Z}_3$ symmetry of its Dynkin diagram. When combining left and right-moving sectors, the level matching constraint allows NS $-$ to be paired only with itself, the other 3 sectors can be paired in any combination. This gives the unprojected spectrum

Sector	$SO(8)$ **rep.**
(NS $-$, NS $-$)	1
(NS $+$, NS $+$)	$8_v \times 8_v$
(NS $+$, R $-$)	$8_v \times 8'$
(NS $+$, R $+$)	$8_v \times 8$
(R $-$, NS $+$)	$8' \times 8_v$
$\cdots$	$\cdots$.

There are in total 10 sectors in this table and (indepedently of the specific fermion-number-projector), one might imagine building a consistent theory from any combination of them. Clearly, there are 2^{10} possibilities to select some subset of these sectors. But this selection can not be random: We want it

(1) Not to contain a tachyon.

(2) To be modular invariant (i.e. invariant under large diffeomorphisms of the torus, for example under exchange of τ and σ and hence under reinterpretation of the direction of time flow, cf. Fig. 15)

(3) To obey certain mutual consistency rules among the selected vertex operators on the world-sheet. (There should be no leftover phase or branch cut when one operator circles another, cf. Fig. 16. The operator product expansion should close or, in other words, it should not be possible to produce a state in scattering which we have excluded from our selection.)

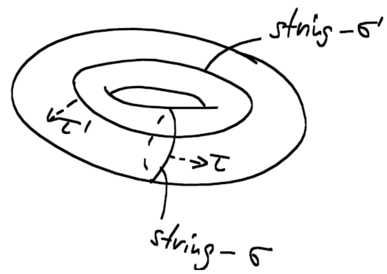

Figure 15: Illustration of the intuitive meaning of modular invariance.

With this, only two inequivalent possibilities of the 2^{10} are left [52]. The corresponding selections are easily formulated using fermion number constraints or projections:

Type IIA: left: $(-1)^F = 1$ right: $(-1)^{\tilde{F}} = 1$ (NS) / $(-1)^{\tilde{F}} = -1$ (R)

Type IIB: left: $(-1)^F = 1$ right: $(-1)^{\tilde{F}} = 1$.

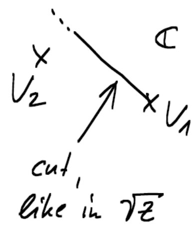

Figure 16: In principle, branch cuts can arise in the correlation function between two vertex operators. This should, however, be forbidden since it makes the integration over all positions impossible.

These general rules translate, specifically in **type IIA** in:

Sector	$SO(8)$ **rep.**		
(NS+, NS+)	$8_v \times 8_v = 1 + 28 + 35$	$=$	$[0]_\phi + [2]_{B_2} + (2)_G$
(NS+, R−)	$8_v \times 8' = 8 + 56'$	$=$	spinor + vector-spinor′
(R+, NS+)	$8 \times 8_v = 8' + 56$	$=$	spinor′ + vector-spinor
(R+, R−)	$8 \times 8' = 8_v + 56_t$	$=$	$[1]_{C_1} + [3]_{C_3}$

To derive the last two columns of this table, one needs elementary representation theory. We will only interpret the results. We note that $SO(8)$ has three inequivalent 56-dimensional representations, two vector-spinors and one antisymmetric rank-2 tensor. We used a square and round bracket for antisymmetric and traceless symmetric tensors of a given rank. Hence e.g. [2] stands for the familiar Kalb-Ramond field and (2) for the graviton. On the bosonic side, we have dilaton, B_2, $g_{\mu\nu}$ and two form-fields, C_1 and C_3. The latter are a new feature of the superstring and the correponding charged states are so-called D0 and D2 branes, which are non-perturbative objects (in the sense that they do not directly follow from the perturbative analysis of world-sheet degrees of freedom). They have to be introduced into the theory for consistency, have their own action and dynamics, and provide potential endpoints for open strings.

We are finding a so-called $\mathcal{N} = 2$ supersymmetric theory since we have two gravitini which are both partners of the same, unique graviton. The other two spinors are known a dilatini. There are two independent SUSY generators and hence SUSY transformations relating the graviton to either one or the other gravitino. However, the overall structure is more involved and all degrees of freedom are needed in fully match fermions and bosons.

Analogously, one finds the field content of the **type IIB** string:

Sector	$SO(8)$ **rep.**		
(NS+, NS+)	$8_v \times 8_v = 1 + 28 + 35$	$=$	$[0]_\phi + [2]_{B_2} + (2)_G$
(NS+, R+)	$8_v \times 8 = 8' + 56$	$=$	spinor′ + vector-spinor
(R+, NS+)	$8 \times 8_v = 8' + 56$	$=$	spinor′ + vector-spinor
(R+, R+)	$8 \times 8 = 1 + 28 + 35_+$	$=$	$[0]_{C_0} + [2]_{C_2} + [4]_{+, C_4}$

The key differences are that this theory is chiral (a preference is given to one of the two different available chiralities of spinors and vector-spinors). Furthermore, the form-field and hence the brane content is different. It is easy to remember that type IIA and IIB theory contain odd and even p-form gauge potentials respectively and hence even and odd Dp-branes. A further noteworthy specialty of the IIB theory is the fact that the C_4 theory is subject to a self-duality constraint, $F_5 = \tilde{F}_5$, which halves the number of degrees of freedom (cf. the index of $[4]_+$ and 35_+).

4.8 Other 10d theories

The name type II refers to the two supersymmetries. There is also a minimally supersymmetric 10d superstring theory called **type I** with unoriented strings. It follows by modding out **worldsheet parity**. By this one means the introduction of an operator Ω which realizes the classical transformation $\sigma \to -\sigma$ at the quantum level (hence $\Omega^2 = 1$) and projecting on the 1-eigenspace of

$$P = \frac{1}{2}(1 + \Omega)\,. \tag{4.69}$$

A detailed analysis reveals that stability ('tadpole-cancellation') always requires the presence of 32 D9-branes, giving rise to gauge fields living in 10d. Due to the projection the group is not $U(32)$ but its 'real subgroup', $SO(32)$.

Furthermore, it is consistent (and allows for tachyon removal) to supersymmetrize only the left- or right-moving half of the worldsheet theory. For obvious reasons such theories are called heterotic and they come in two types, named after their non-abelian gauge group (which are present in both cases): **heterotic** $E_8 \times E_8$ and **heterotic** $SO(32)$.

Not surprisingly, the latter is related to type I by a so-called **duality**. In this particular case, it is a so-called strong-weak duality saying that type I at weak string coupling is identical to heterotic at strong coupling and vice versa. In fact all of the 5 10d theories above are related to each other by dualities, projections or compactifications (see Fig. 17) and are sometimes referred to collectively as the (perturbative corners of) **M-theory**. One usually includes 11d supergravity in this set, although the fundamental objects there appear to be membranes (specifically **M2-branes**) rather than strings and the theory is much less well understood in the ultraviolet. Occasionally, the name M-theory is also used to refer only to 11d-supergavity rather than to the whole set of theories. It is believed that these 6 theories are the calculable, perturbative corners of a more general and not yet fully understood structure - M-theory as 'defined' by the inner region of the 'amoeba' in Fig. 17.

5 10d actions and compactification

5.1 10d supergravities and Type IIB as an example

The existence of a target space action for each consistent string theory and its fundamental relation to the worldsheet perspective has already been discussed in the context of the bosonic string. All that was said there remains true. In particular, there are always the 10d graviton and B_2 field, coupling to the worldsheet (except in the case of the unoriented type I string,

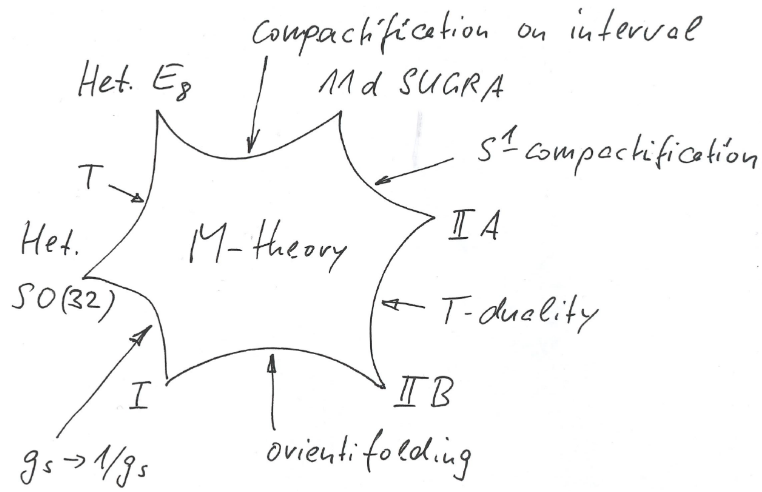

Figure 17: Illustration of M-theory and its perturbative corners: the 5 superstring theories and 11d SUGRA.

where it falls victim to the projection taking us from type IIB to type I). There is also always the dilaton governing the convergence of perturbation theory. Together, dilaton, graviton and B_2 form the NS-NS sector (see above). As a novelty, one has the C_{p+1} (or R-R) form fields and the corresponding Dp-branes. These are dynamical objects, just like the string itself, but with different dimensionality and (at weak string coupling) larger tension. This analysis would be slightly different in the heterotic case, where there are no C-forms but rather gauge fields. Crucially, there are fermionic partners for all the fields above.

What is very different from the bosonic case is the uniqueness of the above 5 theories, independently of the stringy construction. This is due to supergravity. Indeed, supersymmetry (and supergravity) exists in various dimensions (cf. Appendix of volume II of [52]), but its realization becomes harder and harder as the dimension grows. This can be roughly understood by noting that the spinor dimension grows exponentially with D, making it more difficult to find a matching bosonic structure.

Let us start by noting that for even D one has

$$\Gamma_0, \cdots, \Gamma_{D-1} \qquad \text{and} \qquad \Gamma \equiv \Gamma_0 \cdot \Gamma_1 \cdot \cdots \cdot \Gamma_{D-1}\,, \tag{5.1}$$

allowing us to define chirality through the projector $(1+\Gamma)/2$. In the dimension $(D+1)$, which is now odd, Γ becomes the highest 'usual' gamma matrix and the product of all gamma matrices becomes $\Gamma \cdot \Gamma \sim \mathbb{1}$. Hence, chirality can not be defined.

In some dimensions (see [52]) there is a Majorana spinor and, if both Weyl and Majorana spinors exist independently, it is sometimes possible to impose both constraints together. We have seen that this happens in $D = 2$, where the naive spinor dimension is $2^{D/2} = 2$, i.e. 4 real d.o.f., and we found spinors with one real component.

This situation occurs again in $D = 10$, where the Dirac spinor has 32 components and a 16-component real spinor exists. This spinor has 4 times the degrees of freedom of a minimal 4d spinor, hence the minimal 10d SUSY is referrred to as $\mathcal{N} = 4$ SUSY in 4d language. One may also characterize this as 10d $\mathcal{N}=1$ SUSY. We have encountered a gauge theory with this amount of supersymmetry when we quantized the open superstring in 10d. This gauge theory (more precisely its non-abelian version) can be coupled to supergravity and it is the SUSY of the heterotic and type I theories. It is also possible to have 10d $\mathcal{N}=2$ supergravity (correspnding to $\mathcal{N}=8$ in 4d language. Gauge fields can not be added to such theories. This is the SUSY of the type II string. Very interestingly, this higher-dimensional, highly supersymmetric setting is so constraining that (under very reasonable assumptions) these 4 theories can be shown to be **the only 10d supergravities**. This uniqueness includes the gauge group – only $E_8 \times E_8$ and $SO(32)$ are possible on anomaly cancellation grounds. It is very intriguing that precisely these 4 10d supersymmetric field theories are realized in string theory. All these theories are unique also in the sense that no dimensionless parameters are present. An equally unique supergravity theory exists in 11d - it is the 11d theory linked to type IIA via compactification on S^1 as notes earlier. No other supergravity theories above 9 dimensions are known.

All of them have been tried as starting points for a stringy description of the real world. The landscape, i.e. a very large number of potentially suitable 4d models has been most convincingly established in type IIB (although there are still reasonable doubts, to which we will come). We hence focus on this theory.

In the widely used conventions of [52, 63], the bosonic part of the string-frame **type IIB** action reads

$$S=\frac{1}{2\kappa_{10}^2}\int d^{10}x\sqrt{-g}\left\{e^{-2\phi}\left[\mathcal{R}+4(\partial\phi)^2-\frac{1}{2\cdot 3!}H_3^2\right]-\frac{1}{2}F_1^2-\frac{1}{2\cdot 3!}\tilde{F}_3^2-\frac{1}{4\cdot 5!}\tilde{F}_5^2\right\}+S_{CS}+S_{loc}\,. \tag{5.2}$$

Here $2\kappa_{10}^2=(2\pi)^7\alpha'^4$ and

$$\tilde{F}_3=F_3-C_0\wedge H_3\;,\qquad \tilde{F}_5=F_5-\frac{1}{2}C_2\wedge H_3+\frac{1}{2}B_2\wedge F_3\,. \tag{5.3}$$

The RR-form field strengths with a tilde are gauge invariant (as is H_3). This implies special gauge transformation properties of some of the potentials, e.g.

$$C_2\to C_2+d\lambda_1\qquad\text{goes together with}\qquad C_4\to C_4+\frac{1}{2}\lambda_1\wedge H_3\,. \tag{5.4}$$

Furthermore, terms which do not involve the metric are often referred to as Chern-Simons terms. In our case it reads

$$S_{CS}=\frac{1}{4\kappa_{10}^2}\int e^{\phi}\,C_4\wedge H_3\wedge F_3\,. \tag{5.5}$$

Finally, we collect the actions of the various branes (including extended fundamental strings) which may be present in the target space and are described by the 'localized' part S_{loc}. We just display the example of a D3-brane

$$S_{loc}\supset S_{D3}=\frac{1}{2\pi^3\alpha'^2}\int_{D3}C_4\;-\;\int_{D3}d^4\xi\,\sqrt{-g}\,T_3\qquad\text{with}\qquad T_3=\frac{1}{(2\pi)^3\alpha'^2}\,. \tag{5.6}$$

The first part of S_{D3} may also be called a Chern-Simons-type term since the metric does not enter. The coordinates ξ parameterize the world-volume of the brane and the metric next to them is the pullback of the 10d metric. Analogous expressions for the other odd-dimensional

Dp-branes and the string have to be added. The general formula for the tension appearing in the S_{Dp} is

$$T_p = \frac{e^{(p-3)\phi/4}}{(2\pi)^p \alpha'^{(p+1)/2}}\,. \tag{5.7}$$

We do not display the completely analogous expression for type IIA, where the relevant RR form fields are C_1 and C_3. We only note that the non-localized CS term takes the form

$$\int B_2 \wedge F_4 \wedge F_4\,. \tag{5.8}$$

The action for type I follows from that of type IIB upon a so-called orientifold projection, i.e., a projection on states invariant under worldsheet-parity inversion. In 10d, this implies the removal of C_0, B_2 and C_4. Furthermore, 32 D9-branes have to added, also subject to a certain projection, which restricts the gauge group to SO(32). Thus, one basically includes the lagrangian of the corresponding 10d super-Yang-Mills (SYM) theory.

Finally, in the heterotic case one removes the C-forms (keeping B_2) and adds SYM lagrangians with groups $E_8 \times E_8$ or $SO(32)$. It is then clear that the advertised duality between the type-I and the heterotic $SO(32)$ theory also involves the exchange of the F_3 and H_3.

We recall again that the fermionic parts also differ strongly between the various theories, given in particular that SUSY is reduced to 10d $\mathcal{N} = 1$ in all but the two type II theories.

5.2 Kaluza-Klein compactification

One has thus arrived at a possibly fundamental and (involving the various dualities above) unique 10d theory. To describe a 4d world on this basis, the logical procedure is to employ the idea of Kaluza-Klein compactification. This method of obtaining lower from higher-dimensional theories is old and has, as we will see, some appeal in its own right [64–69].

Let us start with what may be the simplest example: a 5d scalar field on $M = \mathbb{R}^4 \times S^1$, where the S^1 has radius R (such that $x^5 \in (0, 2\pi R)$):

$$S = \int_M d^5x\, \frac{1}{2}(\partial_M \phi)(\partial^M \phi)\,, \qquad M \in \{0,1,2,3,5\}\,. \tag{5.9}$$

We take $\phi = 0$ (in fact any other value, $\phi =$const., would be equally good) as our vacuum and parameterize fluctuations around this solution according to

$$\phi(x,y) = \sum_{n=0}^{\infty} \phi_n^c(x)\cos(ny/R) + \sum_{n=1}^{\infty} \phi_n^s(x)\sin(ny/R)\,. \tag{5.10}$$

Here we renamed x^5 according to $x^5 \to y$ and we use the argument x as $x = \{x^0, x^1, x^2, x^3\}$. One immediately finds

$$S = 2\pi R \int d^4x \left[\frac{1}{2}(\partial\phi_0^c)^2 + \frac{1}{4}\sum_{n=1}^{\infty}\left\{(\partial\phi_n^c)^2 + m_n^2(\phi_n^c)^2 + (\partial\phi_n^s)^2 + m_n^2(\phi_n^s)^2\right\}\right]\,, \tag{5.11}$$

with $m_n = n/R$. Hence, our model is exactly equivalent to a 4d theory with one massless field and a (doubly degenerate) tower of KK modes. The massless mode parameterizes a 'flat direction', i.e. it is not only massless but has no potential at all. It can hence take an arbitrary constant value, which would still be a solution. Such a field is called a **modulus**.

We will frequently encounter cases where the value of the modulus governs the masses and couplings of the rest of the 4d theory. To create such a situation in our toy model, enrich our theory by 5d fermions and introduce the 5d coupling

$$\lambda \phi \overline{\psi} \psi \, . \tag{5.12}$$

It is an easy exercise to derive the 4d action as above and read off explicitly how the 4d fermion masses depend on the VEV of ϕ. Now ϕ is more like one of the moduli we will encounter in more realistic cases below. We note that our 'modulus' has the problem that loop corrections will give it a potential even in the 5d local lagrangian. In this sense it is really not a proper modulus. We will see better examples below.

Indeed, let us now turn to the historical example which is most directly associated with the word Kaluza-Klein theory. Consider pure general relativity in 5d,

$$S = \frac{M_{P,5}^3}{2} \int d^4x \, dy \, \sqrt{-g_5} \, \mathcal{R}_5 \, , \tag{5.13}$$

parameterize the metric as

$$(g_5)_{MN} = \begin{pmatrix} g_{\mu\nu} + (2/M_P^2)\phi^2 A_\mu A_\nu & (\sqrt{2}/M_P)\phi^2 A_\mu \\ (\sqrt{2}/M_P)\phi^2 A_\nu & \phi^2 \end{pmatrix} \tag{5.14}$$

where

$$M, N, \dots \in \{0,1,2,3,5\} \qquad \text{and} \qquad \mu, \nu, \dots \in \{0,1,2,3\} \, , \tag{5.15}$$

and we write $M_{P,5}$ and M_P for the 5d and 4d reduced Planck mass respectively.

As above, we assume that $y \in (0, 2\pi R)$ parameterizes an S^1 and we base our analysis on the solution $g_{\mu\nu} = \eta_{\mu\nu}$, $A_\mu = 0$ and $\phi^2 = g_{55} = 1$. Based on our scalar-field example, we expect that the Fourier decomposition of all fields as functions of y will give a 4d theory with a tower of massive modes. Focussing on the zero-modes only corresponds to assuming that all fields are independent of y. Under this assumption, it is straightforward to work out the higher-dimensional action, i.e. the 5d Ricci scalar, in terms of the ansatz (5.14). The result reads [67]

$$S = \int d^4x \, \sqrt{-g} \, \phi \, \left(\frac{M_P^2}{2} \, \mathcal{R} - \frac{1}{4} \phi^2 F_{\mu\nu} F^{\mu\nu} + \frac{M_P^2}{3} \, \frac{(\partial\phi)^2}{\phi^2} \right) \, , \tag{5.16}$$

which can of course be brought to the Einstein frame by $g_{\mu\nu} \to g_{\mu\nu}/\phi$.

The key lessons are that the (zero modes of the) 5d metric degrees of freedom have turned into the 4d metric, an abelian gauge field and a scalar. The appearance of a $U(1)$ gauge theory is not surprising since our starting point, the $\mathbb{R}^4 \times S^1$ geometry, clearly has a global $U(1)$ symmetry. But, since we are in general relativity and our starting point is diffeomorphism invariant, we are also allowed to rotate the S^1 (i.e. shift y) differently at every point x. Hence, our symmetry must actually be a $U(1)$ gauge symmetry.

Moreover, we have a solution of the 5d Einstein equations for every fixed radius R. Thus, we expect a scalar degree of freedom, corresponding to R, with an exactly flat potential. This degree of freedom is the scalar field ϕ. Note that it is sometimes convenient to parameterize the S^1 by the dimensionless variable $y \in (0,1)$ and correspondingly to have $\phi = \sqrt{g_{55}} = 2\pi R$ in the vacuum. We note that, while M_P was originally introduced as a parameter in the metric ansatz, the result (5.16) used the identification

$$M_P^2 = 2\pi R\, M_{P,5}^3\,. \tag{5.17}$$

Before closing this generic Kaluza-Klein section, it will be useful to consider yet another example: Let the geometry again be $\mathbb{R}^4 \times S^1$ and let the 5d lagrangian contain a $U(1)$ gauge theory. For simplicity, we will focus on the Kaluza-Klein or dimensional reduction of this $U(1)$, ignoring the 5d gravity part which we just discussed. Thus, we start with

$$S = \int d^4x\, dy \left(-\frac{1}{4g_5^2}\, F_{MN}F^{MN} \right)\,, \tag{5.18}$$

where $g_{MN} = \eta_{MN}$ and $y \in (0, 2\pi R)$. With the ansatz

$$A_M = (A_\mu, \phi) \tag{5.19}$$

one finds, at the zero-mode level,

$$S = \int d^4x \left(-\frac{1}{4g^2}\, F_{\mu\nu}F^{\mu\nu} - \frac{1}{2g^2}(\partial\phi)^2 \right)\,. \tag{5.20}$$

The crucial lesson is that the 5d gauge field gives rise to a 4d gauge field and a scalar, the latter being associated to A_5 or, in a more geometrical language, to the **Wilson line** integral

$$\oint A = \int dy\, A_5 = 2\pi R\, \phi(x)\,. \tag{5.21}$$

This Wilson line measures the phase which a charged particle acquires upon moving once around the S^1, just as in the Aharonov-Bohm experiment. Assuming that the minimally charged particle (we do not display the corresponding part of the lagrangian) has unit charge, the phase measured by ϕ becomes equivalent to zero for $\phi = 1/R$. Thus, we have found an exactly massless (at the classical level) periodic scalar field, also known as an **axion** or **axion-like particle** or **ALP**.

Let us draw a lesson from the above which will also be important for string compactifications, to be discussed below: We have seen two types of moduli arise, one associated to the geometry of the compact space (g_{55}), the other to the gauge field configuration in the compact space (A_5). Both have no classical potential since in one case 5d diffeomorphism invariance, in the other case 5d gauge invariance forbid the corresponding potential term. Moreover, due to this symmetry argument 5d loop corrections do not induce such a potential. However, in both cases 4d loop corrections can provide a potential and hence give a mass to the above fields. This is not in contradiction to the symmetry argument just stated since 4d loop effects can in general not be written in terms of 5d local operators. However, in the presence of enough supersymmetry in the resulting 4d theory, these loop corrections may be forbidden such that the relevant moduli remain exactly massless or more precisely, their potential remains identically zero as an exact statement. This generally happens in 4d $\mathcal{N} = 2$ SUSY.

5.3 Towards Calabi-Yaus

We now want to explain how the 10d SUGRA theories provided by the superstring can be compactified to 4d. There are two approaches: We could start by developing the toy model path started in the previous section, i.e., we could consider the geometry $\mathbb{R}^9 \times S^1$. This would give us a 9d theory, without too many new features (except for supersymmetry, which would keep all moduli massless). Next, we could consider $\mathbb{R}^8 \times T^2$. We would now encounter moduli corresponding to g_{88}, g_{99} and g_{89}, characterizing both size and the shape of the torus. Thus, we would get an 8d theory with (at least) 3 scalars corresponding to geometric moduli. Much more could be said about compactifications on tori and related simple geometries.

However, we will take a different approach and first introduce a much more general and powerful set of examples - the Calabi-Yau geometries. These are the compactification spaces on which the landscape as we presently understand it is mostly built. Later on, we will return to tori to illustrate some of the more abstract concepts used.

Our key starting point is the desire to find a **solution** of the 10d equations of motion corresponding to a 4d world. Setting all fields execpt the metric to zero, this implies that we must have $(\mathcal{R}_{10})_{MN} = 0$ to solve Einstein's equations. This condition is called **Ricci flatness** and it is obviously satisfied for S^1 and (flat) tori T^n mentioned above. The interesting and non-trivial fact is that there exists a large class of relatively complicated compact 6d manifolds which are also Ricci flat and hence represent suitable compactification spaces: **the Calabi-Yau manifolds**.

Before giving the definition, we need a few geometrical concepts. Our treatment will be extremely brief and hence, unfortunately, superficial. Much more material can be found in e.g. in [51, 53, 70–72].

To begin, Calabi-Yaus are complex manifolds. This is a fairly straightforward generalization of the familiar concept of a 2n-dimensional real differentiable manifold X. The key new point is that the charts

$$(U_i, \phi_i)\ , \qquad \phi_i : U_i \to \phi_i(U_i) \subset \mathbb{C}^n\ , \tag{5.22}$$

are now maps from open sets U_i of X to $\mathbb{C}^n$, with they key compatibility condition being that the functions $\phi_j \circ \phi_i^{-1}$ are holomorphic. In other words, our manifold locally looks like $\mathbb{C}^n$ and coordinate changes are of the form

$$z'^i = z'^i(z^1, \cdots, z^n)\,, \tag{5.23}$$

with any appearance of $\overline{z}^{\overline{\imath}}$ in the argument of the new coordinate excluded.

On a complex manifold, it makes sense to complexify tangent and cotangent space as well as all their higher tensor products. Thus, tensor fields are complex. For example, local bases of tangent and cotangent space are provided by

$$\frac{\partial}{\partial z^i}\,,\ \frac{\partial}{\partial \overline{z}^{\overline{\imath}}} \qquad \text{and} \qquad dz^i\,,\ d\overline{z}^{\overline{\imath}}\,, \tag{5.24}$$

with $z^i = x^i + iy^i$ etc. It is natural to define the tensor

$$J = i\, dz^i \otimes \frac{\partial}{\partial z^i} - i\, d\overline{z}^{\overline{\imath}} \otimes \frac{\partial}{\partial \overline{z}^{\overline{\imath}}}\,, \tag{5.25}$$

which roughly speaking corresponds to 'multiplication by i' in cotangent space. Its components are

$$J = \begin{pmatrix} i\mathbb{1} & 0 \\ 0 & -i\mathbb{1} \end{pmatrix} \tag{5.26}$$

in a complex basis and

$$J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \tag{5.27}$$

in a real basis. A crucial feature is $J^2 = -\mathbb{1}$.

A real manifold with a tensor J as above is called an almost complex manifold and J is called an almost complex structure. If such a J satisfies a certain integrability condition (vanishing of the Niejenhuis tensor), complex coordinates can be given and J turns into the so-called complex structure of a complex manifold. We will only be interested in this latter case.

Even more, we will demand that our manifold has a metric (is a Riemannian manifold) and that this metric is compatible with J. In other words, we demand that J is covariantly constant. This turns the manifold into a Kahler manifold, a concept which we already used when discussing field spaces of supersymmetric field theories. We will not demonstrate this but give right away a stronger definition: A complex manifold with a metric is called Kahler if the metric can locally be written as

$$g_{i\bar{\jmath}} = \frac{\partial^2 K}{\partial z^i \, \partial \overline{z}^{\bar{\jmath}}}\,, \tag{5.28}$$

with K a real function defined in every patch and with $g_{ij} = g_{\bar{\imath}\bar{\jmath}} = 0$. We note that this last condition by itself would make the metric hermitian, but we are interested only in the stronger Kahler condition.

We also note that the metric allows us to lower the second index of J, turning J into a rank-2 lower-index tensor. This tensor turns out to be antisymmetric and hence defines a 2-form, the so-called **Kahler form**

$$J = i g_{i\bar{\jmath}} \, dz^i \wedge d\overline{z}^{\bar{\jmath}}\,. \tag{5.29}$$

We see that, given a complex structure, the 2-form J determines the metric and vice versa. This will become important below when we will be discussing different metrics on the same differentiable manifold.

Next, we need the concept of **holonomy**. We know form basic differential geometry that, with a metric, one gets a unique Riemannian or Levi-Civita connection and hence the possibility to parallel-transport tangent vectors. Given any point $p \in X$ and any closed curve C beginning and ending in p, we hence have a linear map

$$R(C) : T_p \to T_p \qquad \text{or} \qquad R(C) \in SO(2n)\,. \tag{5.30}$$

The latter statement follows if we assume orientability (for complex manifolds this is guaranteed) and recall that the Riemannian parallel transport does not change the length of a vector. It can be shown that the set of all $R(C)$ forms a group and that this group does not depend on the choice of p (assuming X is connected). This is the holomomy group.

We are now in the position to give one (of the many equivalent) definitions of a Calabi-Yau: **A Calabi-Yau 3-fold (our case of interest) is a compact, complex Kahler manifold**

with $SU(3)$ holomomy. More generally, for a complex n-fold one demands that the holonomy is $SU(n) \subset SO(2n)$. As we will argue in a moment, this implies that some of the 10d supersymmetry is preserved in the 4d effective field theory and that Einstein equations are solved without sources (Ricci flatness).

Though the Einstein equations are maybe physically more important, we will start with SUSY. Very superficially, we expect that a 4d supersymmetric effective theory will have massless spinors. Hence spinors need to have zero-modes which, in the simplest case, corresponds to the existence of covariantly constant spinors on the compactification space. We will see in a moment that this covariantly constant spinor is intimately linked to $SU(3)$ holonomy.

But let us first give a more careful argument for why unbroken SUSY requires the compact space to have a covariantly constant spinor: While we have not given the supergravity transformations of the various fields in 10d, we may recall the 2d case of worldsheet-SUGRA: Here, we have seen that the transformation of the gravitino is proportional to the covariant derivative of the SUSY parameter, i.e. of the spinor $\xi(\sigma)$:

$$\delta_\xi \chi_a = \nabla_a \xi \,. \tag{5.31}$$

This is similar in 10d. Hence, to identify a 4d SUSY parameter under which the vacuum is invariant, one needs a covariantly constant spinor. On a curved manifold this is a non-trivial requirement.

To see this in more detail, we need the group-theoretic fact that $SO(6) = Spin(6)/\mathbb{Z}_2$, $Spin(6) = SU(4)$ and that the 6d Weyl spinor transforms in the $\mathbf{4}$ of $SU(4)$, using this isomorphism. Furthermore, we embed our 10d spinor in the tensor product of 4d spinor and 6d spinor. Since 4d space is flat, the critical issue for the constancy of our 10d spinor is the constancy of its 6d spinor part. In other words, we have to take the 6d spinor to be covariantly constant along X. Furthermore, without loss of generality we assume that this spinor takes the form

$$\xi(p) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \xi_0(p) \end{pmatrix} \tag{5.32}$$

at some point $p \in X$. Since it is part of a covariantly constant spinor field, the parallel transort will follow this field and, in particular, bring $\xi(p)$ back to itself for any loop C. But this clearly means that the holonomy matrices may only act on the first 3 components, i.e. we need $SU(3)$ holonomy.

The reverse is obvious: Given $SU(3)$ holomomy, a covariantly constant spinor can be constructed by parallel transporting $\xi(p)$ given above to any point of X. The only way in which this might fail is if the construction were ambiguous, i.e., if two different paths from p to p' gave rise to two different spinors $\xi(p')$. But this would imply that a closed path starting at p exists along which the parallel transport of $\xi(p)$ is non-trivial. This would be in contradiction to $SU(3)$ holonomy. Thus, we have the equivalence between $SU(3)$ holonomy and the existence of a covariantly constant spinor, i.e. the survival of 4d SUSY.[10]

[10]More precisely, 4d $\mathcal{N} = 2$ SUSY in the type II case and 4d $\mathcal{N} = 1$ SUSY in the type I and heterotic case. The reason is the presence of two independent 10d SUSY generators in the former situation.

Next, we consider Ricci flatness. We first note that, on Kahler manifolds, the only non-zero components of the Levi-Civita connection are

$$\Gamma_{ij}{}^{k} = g^{k\bar{l}}\partial_i g_{j\bar{l}} \qquad \text{and} \qquad \Gamma_{\bar{\imath}\bar{\jmath}}{}^{\bar{k}} = g^{\bar{k}l}\partial_{\bar{\imath}} g_{l\bar{\jmath}}\,. \tag{5.33}$$

This leads to significant simplifications for the Riemann tensor and the Ricci tensor which we do not work out. For example, the only non-vanishing Riemann tensor components are of the form

$$R_{i\bar{\jmath}k\bar{l}} \tag{5.34}$$

and those related by antisymmetry in the first and second index pair. In other words, the first two and last two indices have to be of opposite type (holomorphic and antiholomorphic). Moreover, the Ricci tensor can be written as

$$R_{i\bar{\jmath}} = \partial_i \partial_{\bar{\jmath}} \ln \det g\ . \tag{5.35}$$

As is well known, the significance of $R_{i\bar{\jmath}}{}^{\alpha}{}_{\beta}$ is that, if intepreted as a matrix with indices α and β, it describes the rotation of a covector upon parallel transport along a loop with orientation specified by i and $\bar{\jmath}$. Here we use greek letters for the second index pair to emphasize that they can take either holomorphic or antiholomorphic values, e.g. $\alpha = k$ or $\alpha = \bar{k}$. The previously noted restrictions on holomorphy/antiholomorphy of the second index pair means that either $(\alpha, \beta) = (k, l)$ or $(\alpha, \beta) = (\bar{k}, \bar{l})$. This can straightforwardly be shown to imply that the corresponding rotation matrix is in the $U(n)$ subgroup of the general holonomy group $SO(2n)$. More generally, the conditions of a manifold being Kahler and having $U(n)$ holonomy are equivalent.

Since $U(n) = SU(n) \times U(1)$, the spin connection of Kahler manifolds can be thought of as the sum of an $SU(n)$ and a $U(1)$ connection. The latter is just a standard $U(1)$ connection, like in the case of an abelian gauge theory. Its field strength $F_{i\bar{\jmath}}$ being non-zero characterizes the holonomy not being restricted to $SU(n)$.

Concretely, recall that the complex structure is defined as multiplication by 'i' on the cotangent or tangent vector space. In components, the corresponding operator or matrix is $J^{\alpha}{}_{\beta}$, which is hence the generator of the $U(1)$. The $U(1)$ part of the $U(n)$ field strength encoded in $R_{i\bar{\jmath}}{}^{\alpha}{}_{\beta}$ can hence be determined from the projection on $J^{\alpha}{}_{\beta}$. An explicit definition is

$$F_{i\bar{\jmath}} \equiv \mathrm{tr}[\tilde{R}_{i\bar{\jmath}}J] \equiv R_{i\bar{\jmath}}{}^{\alpha}{}_{\beta}J^{\beta}{}_{\alpha} = iR_{i\bar{\jmath}}{}^{k}{}_{k} - iR_{i\bar{\jmath}}{}^{\bar{k}}{}_{\bar{k}} = 2iR_{i\bar{\jmath}}{}^{k}{}_{k} = -2iR_{i}{}^{k}{}_{\bar{\jmath}k} = -2iR_{i\bar{\jmath}}\,. \tag{5.36}$$

Here the $\tilde{R}$ denotes the Riemann tensor with suppressed second index pair, not the Ricci tensor. The final manipulations leading to the Ricci tensor require the use of the symmetry properties of the Riemann tensor together with the Kahler property of our manifold. We leave that as a problem (see e.g. [71]). Eventually, we see that the $U(1)$ field strength components equal those of the Ricci tensor up to a prefactor (note however the different symmetry properties of the two tensors). Thus, $SU(n)$ holonomy is equivalent to Ricci flatness.

A final important point concerns the definition of Calabi-Yaus via the Chern class. Note first that, due to the $U(n)$ holonomy (or equivalently because of the special index structure of the Riemann tensor), the tangent bundle of Kahler manifolds can be viewed as a complex vector bundle with the curvature specified by $R_{i\bar{\jmath}}{}^{k}{}_{l}$. In other words, one can consider the curvature 2-form

$$R(T_X) = dz^i \wedge d\bar{z}^{\bar{\jmath}}\, R_{i\bar{\jmath}}{}^{k}{}_{l}\,, \tag{5.37}$$

which takes its values in $Lie(U(n))$. It is possible to write down the multi-form

$$c(X) = \det(\mathbb{1} + R(T_X))\,, \tag{5.38}$$

where the determinant refers to the matrix indices and multiplication relies on the wedge product. It is then expanded according to

$$c(X) = 1+c_1(X)+c_2(X)+\cdots = 1+\mathrm{tr}\,R(T_X)+\mathrm{tr}\,\big(R(T_X)\wedge R(T_X) - 2(\mathrm{tr}\,R(T_X)^2)\big)+\cdots\,. \tag{5.39}$$

Here $c_k(X)$ is a $(2k)$-form, defining the k th **Chern class**. Concretely, the 1st Chern class is said to be zero if c_1 is **exact**, which means that $c_1 = d\omega$ for some ω. More formally, this means that c_1 is zero in **cohomology**, a concept we will discuss next. Crucially, while c_1 was defined using the metric, it is invariant (up to exact pieces) under smooth variations of the latter. It hence represents a **topological invariant**.

After these preliminaries, we can formulate the celebrated theorem by **Yau**: **Let X be a Kahler manifold and J its Kahler form. If the 1st Chern class vanishes, then a Ricci flat metric with Kahler form J' in the same cohomology class can be given. This so-called Clabi-Yau metric is unique.**

Being in the same cohomology class means that $J - J'$ is exact. The key point is that, in practice, finding the Calabi-Yau metric is very hard (it has not been achieved analytically in any example). However, checking the topological condition of vanishing 1st Chern class is easy and guarantees the existence of many (explicitly known) suitable complex manifolds on which we hence know that a Calabi-Yau metric exists. But one will in general not find the metric for which c_1 is zero as a 2-form, only one with $c_1 = d\omega$.

5.4 Homology and cohomogy

We are overdue with developing a few more simple mathematical ideas concerning in particular differential forms and topology. We start with homology and define a **p-chain** as the formal sum of over p-dimensional submanifolds S_i of some compact manifold X:

$$c_p = \sum_i \alpha_i\,(S_p)_i\,. \tag{5.40}$$

Depending on whether the coefficients α_i are real, complex, integer etc. one can be talking about holomolgy over the real, complex or integer numbers. In the first two cases, the p-chains form real and complex vector spaces respectively.

One can consider the boundary of each $(S_p)_i$ and hence of c_p, which is a $(p-1)$-dimensional submanifold. Taking the boundary is denoted by the doundary operator ∂. A chain without boundary,

$$\partial c_p = 0\,, \tag{5.41}$$

is called a cycle. Crucially, ∂^2 is zero, in other words, a boundary has itself no boundary. A few simple examples are given in Fig. 18.

Given the linear operator ∂ with $\partial^2 = 0$, it is natural to consider its homology groups:

$$H_p = \frac{\mathrm{Ker}(\partial_p)}{\mathrm{Im}(\partial_{p+1})} = \frac{p\text{-cycles}}{\text{boundaries of } (p+1)\text{-chains}}\,. \tag{5.42}$$

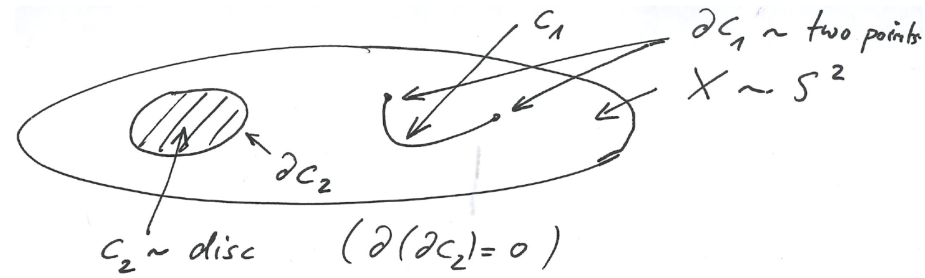

Figure 18: Some simple submanifolds and their boundaries.

The word group refers to addition, in the sense in which every vector space is an abelian group. The index p of ∂_p denotes the restriction of ∂ to the space of p-chains. We will suppress this index when it is clear from the context on which objects ∂ acts. As an example, we display certain 1-cycles on the genus-2 Riemann surface R_2 in Fig. 19. It is easy to convince oneself that, working over the real numbers, $H_1(R_2)$ is 4-dimensional. Representatives a, b, c and d of the 4 linearly independent **homology classes** (the elements of H_p) are shown.

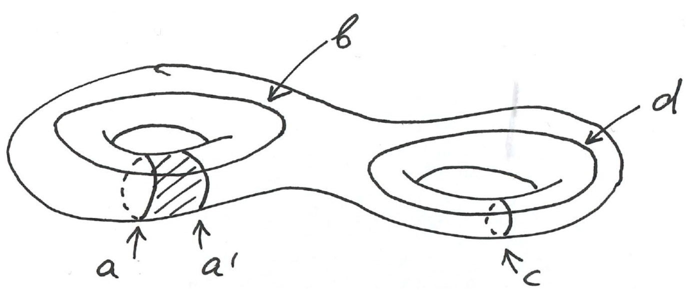

Figure 19: Representatives of the 4 linearly independent homology classes in $H_1(R_2)$. The cycles a and a' are equivelent since they differ by a boundary - the boundary of the hatched 2-dimensional submanifold.

As another example, consider the 3-torus T^3 and convince yourself (at the intuitve level) that $\dim(H_0) = 1$ (which corresponds to T^3 being connected), $\dim(H_1) = \dim(H_2) = 3$ and $\dim(H_3) = 1$. If the torus is thought of as $\mathbb{R}^3$ modulo discrete translations, representatives of H_2 can be thought of as 3 planes, each orthogonal to one of the three axes.

One calls the above **simplicial homology**.

Now we turn to p-forms as the 'dual' objects with respect to the chains. So far, we use the word duality at an informal level, meaning simply that at chain c_p and a form ω_p can be combined in a natural way to give a number:

$$\omega_p(c_p) = \int_{c_p} \omega_p = \sum_i \alpha_i \int_{(S_p)_i} \omega_p \,. \tag{5.43}$$

On the space of forms, we have an operator analogous to ∂, which also also squares to zero: It is the exterior derivative d or, restricted to p-forms, d_p:

$$d_p : \omega_p \to \omega_{p+1} = d\omega_p \,. \tag{5.44}$$

Thus, it is natural to consider the cohomology groups of the **de Rham cohomology**:

$$H^p = \frac{\mathrm{Ker}(d_p)}{\mathrm{Im}(d_{p-1})} = \frac{\text{closed } p\text{-forms}}{\text{exact } p\text{-forms}} \,. \tag{5.45}$$

In the last expression, we use the definition that a p-form ω_p is called **closed** if $d\omega_p = 0$. Similarly, it is called **exact** if it can be written as $\omega_p = d\omega_{p-1}$.

It is easy to convince oneself that the pairing (5.43) between chains c_p and forms ω_p induces a pairing between the correspsonding cohomology classes, sometimes denoted by $[c_p]$ and $[\omega_p]$. In other words,

$$\int_{c_p} \omega_p \tag{5.46}$$

is independent of the representative. For example, one has

$$\int_{c_p} (\omega_p + d\omega_{p-1}) = \int_{c_p} \omega_p + \int_{\partial c_p} \omega_{p-1} = \int_{c_p} \omega_p \,, \tag{5.47}$$

since c_p has no boundary. Analogously, replacing c_p by $c_p + \partial c_{p+1}$ does not affect the pairing.

One can furthermore show that this pairing between homology and cohomology classes is not degenerate and that hence

$$H_p(X) = H^p(X)^* \,, \tag{5.48}$$

i.e. they are dual vector spaces (cf. de Rham's theorems). In particular, their dimensions coincide, defining the so-called **Betti numbers**

$$b_p(X) = \dim H_p(X) = \dim H^p(X) \tag{5.49}$$

of the manifold X. Intuitively, they coint the number of inequivalent basis p-cycles.

We note that, if $\dim X = n$, there also exists a natural pairing between p-cycles and $(n-p)$-cycles: the intersection number. For example, given T^3, a 1-cycle (a line) and a 2-cyle (a plane), one can find out whether the two intersect (intersection number one) or don't (intersection number zero). For a Riemann surface, the pairing is between 1-cycles and 1-cycles, with the meaning of the intersection number being obvious from Fig. 19. It is intuitively clear that this lifts to a paring between homology classes.

The analogue of this on the cohomology side is

$$[\omega_p] \cdot [\omega_{n-p}] = \int \omega_p \wedge \omega_{n-p} \,. \tag{5.50}$$

This pairing is also non-degenerate and hence turns H^p into the dual of the vector space H^{n-p}. But since we already know that H_{n-p} is the dual of H^{n-p}, we have found a canonical isomorphism

$$H^p(X) \cong H_{n-p} \,. \tag{5.51}$$

This is known as **Poincaré duality**. To say this more explicitly, a p-form ω_p is Poincare dual to an $(n-p)$-cycle c_{n-p} if

$$\int_{c_{n-p}} \omega_{n-p} = \int \omega_p \wedge \omega_{n-p} \,, \qquad \forall \omega_{n-p} \,. \tag{5.52}$$

More structure arises if a metric is present. In particular, with a metric comes the Hodge star operator,

$$* : \omega_p \;\mapsto\; (*\omega)_{n-p} \qquad \text{with} \qquad (*\omega)_{\mu_{p+1}\cdots\mu_n} = \frac{\sqrt{g}}{p!}\,\omega^{\mu_1\cdots\mu_p}\,\epsilon_{\mu_1\cdots\mu_n}\,. \tag{5.53}$$

This gives rise to scalar product on the space of p-forms,

$$(\omega_p, \alpha_p) = \int_X \omega_p \wedge *\alpha_p\,. \tag{5.54}$$

As a result, one can define the adjoint of d, the so-called co-differential $d^\dagger$. On forms of degree p, it takes the explicit form

$$d^\dagger = (-1)^p *^{-1} d\, *\;. \tag{5.55}$$

With this, one defines the Laplace operator

$$\Delta = d^\dagger d + d\, d^\dagger\,. \tag{5.56}$$

A form is called **harmonic** if $\Delta\omega = 0$. This definition gives rise to the **Hodge decomposition theorem**, which states that on a compact manifold X any form has a *unique* decomposition in an exact, a coexact and a harmonic piece:

$$\omega = d\alpha + d^\dagger\beta + \gamma \qquad \text{with} \qquad \Delta\gamma = 0\,. \tag{5.57}$$

It can furthermore be shown that β vanishes if ω is closed. As a result, any representative of a given cohomology class has a unique decomposition in an exact and harmonic piece. In other words, there is a unique harmonic form in any cohomology class. Intuitively speaking, this is the constant form with the right integral on all cycles (these integrals being fixed by the class). To give a simple concrete example, consider T^2 being parameterized by $(x, y) \in [(0, 1) \times (0, 1)]$. The harmonic one form with integral zero on the x-cycle and integral 1 on the y-cycle is obviously given by $\omega = dy$. A non-harmonic form in the same class would e.g. be $\omega = (1 + \sin(2\pi y))\, dy$.

Finally, it is possible to take the above to the realm of complex manifolds. To do so, recall that on a complex manifold a 1-form takes the form

$$\omega(z, \overline{z}) = \omega(z, \overline{z})_i dz^i + \omega(z, \overline{z})_{\overline{\imath}} d\overline{z}^{\overline{\imath}} \equiv \omega_{(1,0)} + \omega_{(0,1)}\,. \tag{5.58}$$

In other words, we can decompose it in its $(1, 0)$ and $(0, 1)$ parts. The first corresponds to a tensor with one holomorphic and no antiholomorphic index, the second to a tensor with no holomorphic and one antiholomorphic index.

Such a decomposition carries over to higher forms (i.e. antisymmetric tensors) and to cohomology classes. For example,

$$\omega_3 = \omega_{(3,0)} + \omega_{(2,1)} + \omega_{(1,2)} + \omega_{(0,3)}\,, \tag{5.59}$$

where, e.g.,

$$\begin{aligned} \omega_{(2,1)} &= \omega_{ij\overline{k}}\, dz^i \wedge dz^j \wedge d\overline{z}^{\overline{k}} + \omega_{i\overline{\jmath}k}\, dz^i \wedge d\overline{z}^{\overline{\jmath}} \wedge dz^k + \omega_{\overline{\imath}jk}\, d\overline{z}^{\overline{\imath}} \wedge dz^j \wedge dz^k \\ &= 3\, \omega_{ij\overline{k}}\, dz^i \wedge dz^j \wedge d\overline{z}^{\overline{k}}\,. \end{aligned} \tag{5.60}$$

To see the corresponding, refined cohomology construction more expliclitly, recall that the exterior derivative has the particularly compact definition

$$d = dx^a \, \frac{\partial}{\partial x^a} \, . \tag{5.61}$$

Here the partial derivative is supposed to act on the coefficients of any given form and, subsequently, dx^a has to be multiplied with the form using the wedge product from the left. Let us consider specifically a manifold of complex dimension n (real dimension $2n$), such that $a = 1, 2, \cdots, 2n$. Then it is easy to check that

$$d = dz^i \frac{\partial}{\partial z^i} + d\overline{z}^{\bar\imath} \frac{\partial}{\partial \overline{z}^{\bar\imath}} \, , \tag{5.62}$$

or

$$d = \partial + \overline{\partial} \qquad \text{with} \qquad \partial = dz^i \frac{\partial}{\partial z^i} \qquad \text{and} \qquad \overline{\partial} = d\overline{z}^{\bar\imath} \frac{\partial}{\partial \overline{z}^{\bar\imath}} \, . \tag{5.63}$$

Here $i = 1, 2 \cdots, n$. Furthermore, the **holomorphic and antiholomorphic exterior derivatives** square to zero:

$$\partial^2 = \overline{\partial}^2 = 0 \, . \tag{5.64}$$

This permits the construction of a corresponding cohomology, the result being independent of whether ∂ or $\overline{\partial}$ is used. Conventionally $\overline{\partial}$ is used. Thus, one defines the **Dolbeault cohomology**

$$H^{p,q} = \frac{\mathrm{Ker}(\overline{\partial}_{p,q})}{\mathrm{Im}(\overline{\partial}_{p,q-1})} \, , \tag{5.65}$$

which contains finer information than the de Rham cohomology. One may say that it characterizes the interrelation between the non-trivial cycles and the complex structure. We also note the so-called Hodge decomposition

$$H^k = \oplus_{p+q=k} H^{p,q} \, . \tag{5.66}$$

The dimensions of Dolbeault cohomology groups are known as **Hodge numbers**,

$$h^{p,q}(X) \equiv \dim H^{p,q}(X) \, . \tag{5.67}$$

They are commonly arranged in a so-called **Hodge diamond**. With a view to our application to Calabi-Yau manifolds, we display the general form for the case of a complex 3-fold:

$$\begin{array}{ccccccc}
 & & & h^{0,0} & & & \\
 & & h^{1,0} & & h^{0,1} & & \\
 & h^{2,0} & & h^{1,1} & & h^{0,2} & \\
h^{3,0} & & h^{2,1} & & h^{1,2} & & h^{0,3} \\
 & h^{3,1} & & h^{2,2} & & h^{1,3} & \\
 & & h^{3,2} & & h^{2,3} & & \\
 & & & h^{3,3} & & &
\end{array} \, . \tag{5.68}$$

5.5 Calabi-Yau moduli spaces

Due to $SU(3)$ holonomy, the hodge diamond for a Calabi-Yau 3-fold is very special. Using the same arrangement as in (5.68), it reads

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & h^{1,1} & & 0 & \\ 1 & & h^{2,1} & & h^{2,1} & & 1 \\ & 0 & & h^{1,1} & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array} . \tag{5.69}$$

Here, the simplifications arising from the vertical and horizontal reflection symmetry of the Hodge diamond (e.g. $h^{1,1} = h^{2,2}$) are generic - they hold for any complex n-fold. Furthermore, connectedness implies $h^{0,0} = h^{3,3} = 1$. But some features are specific to Calabi-Yaus, such as $h^{1,0} = h^{2,0} = 0$ and, crucially, $h^{3,0} = h^{0,3} = 1$. The latter implies the existence of a unique holomorphic, harmonic 3-form which is convnntionally denotes by Ω,

$$\Omega = \Omega_{ijk}(z)\, dz^i \wedge dz^j \wedge dz^k\,. \tag{5.70}$$

Its existence can be understood on the basis of the covariantly constant spinor ψ:

$$\Omega_{ijk} \sim \overline{\psi}\Gamma_{ijk}\psi\,. \tag{5.71}$$

We will not argue for uniqueness. It however useful to note that the existence of a harmonic, holomorphic 3-form Ω can be used as a defning feature for Calabi-Yau spaces: More generally, a Calabi-Yau n-fold can be defined as a Kahler manifold with a **trivial canonical bundle**. The latter is the nth exterior power of the cotangent bundle - this is the bundle in which Ω lives and which is trivial exactly if there is a nowhere vanishing section - in our case the n-form Ω.

Now, given a Calabi-Yau 3-fold, Yau's theorem guarantees the existence of a unique (given Kahler class and complex structure) Ricci flat metric $g_{i\bar{\jmath}}$. A key question for physics is whether this metric can be defomed maintaining Ricci-flatness since this would imply the existence of moduli:

$$g_{i\bar{\jmath}}\, dz^i\, d\overline{z}^{\bar{\jmath}} \quad\to\quad g_{i\bar{\jmath}}\, dz^i\, d\overline{z}^{\bar{\jmath}} + \delta g_{i\bar{\jmath}}\, dz^i\, d\overline{z}^{\bar{\jmath}} + \delta g_{ij}\, dz^i\, dz^j + \text{h.c.} \tag{5.72}$$

Clearly, if such deformations exist then, not to contradict the uniqueness part of Yau's theorem, they must be accompanied by a change of either the Kahler class or the complex structure. This is indeed the case: A change of the metric of type $\delta g_{i\bar{\jmath}}$ can be directly interpreted as change of (the harmonic reperesentative of) the Kahler form J. The number of such independent deformations, also called **Kahler deformations** is counted by $h^{1,1}$. This number is at least unity since it is always possible to simply rescale the metric, making our manifold larger or smaller without changing its shape.

By contrast, a deformation of type δg_{ij} violates the hermiticity assumption and and it must hence be accompanied by a change of the complex structure if one wants to restore explicitly the Calabi-Yau situation after adding this δg to the original metric. To count these deformations it is useful to define a $(2,1)$ form

$$\delta\chi = \Omega_{ij}{}^{\overline{k}}\, \delta g_{\overline{kl}}\, dz^i \wedge dz^j \wedge d\overline{z}^{\overline{l}} \;\in\; H^{2,1}(X) \tag{5.73}$$

associated with $\delta g_{\overline{kl}}$. Here the index k of Ω_{ijk} has been raised using the Calabi-Yau metric. It can be shown that this represents a one-to-one map between distinct **complex structure deformations** (and hence corresponding metric deformations) and linearly independent Dolbeault cohomology classes of type $(2,1)$. Here by distinct we mean those *not* corresponding to reparametrizations $z^i \to z'^i$.

There is another way of understanding the counting of complex structure deformations: Think of the complexified vector space of 3-cycles, with dimension $2h^{2,1}+2$. Two directions are distinguished by Ω and $\overline{\Omega}$, a feature only visible in Dolbeault but not in de Rham cohomology. Now, the change of complex structure is accompanied by a change of the direction of Ω (and hence of $\overline{\Omega}$) in this space. In other words, Ω is infinitesimally rotated and these possible rotations are parameterized by $h^{2,1}$ complex numbers. One may say that there are $h^{2,1}$ complex directions in which Ω can develop new, infinitesimal components.

One can also invert the equations above, i.e., given a harmonic $(2,1)$-form $\delta\chi$, one can explicitly write down how Ω and the metric change:

$$\delta g_{\bar{\imath}\bar{\jmath}} = -\frac{1}{||\Omega||^2}\,\overline{\Omega}_{\bar{\imath}}{}^{kl}\,\delta\chi_{kl\bar{\jmath}}\,, \qquad \delta\Omega = \delta\chi\,. \tag{5.74}$$

with the constant

$$||\Omega||^2 = \frac{1}{3!}\Omega_{ijk}\overline{\Omega}^{ijk}\,. \tag{5.75}$$

Together with the previously discussed relation

$$\delta g_{i\bar{\jmath}} = -i\delta J_{i\bar{\jmath}}\,, \tag{5.76}$$

we now see explicitly how the cohomology groups $H^{1,1}(X)$ and $H^{2,1}$ (viewed as a subspace of $H^3(X)$, that moves as the complex structure changes) play a central role in describing the moduli space of a Calabi-Yau. An illustration of this has been attempted in Fig. 20. In addition to the textbook literature given earlier, the reader may want to consult [73, 74] for more details.

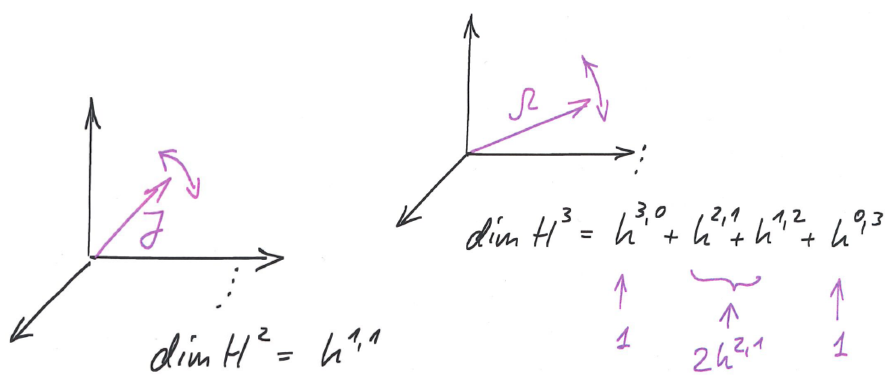

Figure 20: A visuablization attempt of how J and Ω move in the spaces $H^2(X)$ and (the complexification of) $H^3(X)$, thereby determining the metric on a Calabi-Yau.

Before characterizing Calabi-Yau moduli spaces more quantitatively, we want to give at least the simplest example. To do so, let us start with an important set of examples for compact, complex Kahler manifolds: the so-called complex projective spaces. To begin, recall that a **real projective space** $\mathbb{R}P^n$ is $\mathbb{R}^n \setminus \overline{0}$ modulo the equivalence relation $\overline{x} \sim \lambda\overline{x}$ with $\lambda \in \mathbb{R} \setminus 0$. Intuitively speaking, this is the set of lines through the origin, which can easily be given a differentiable structure. For the case of $\mathbb{R}P^2$, the real projective plane, we can equivalently think of $S^2/\mathbb{Z}_2$ – a sphere with antipodal points identified.

This has a natural generalization in the complex numbers: the complex projective space $\mathbb{C}P^n$. They are defined analogously as the set of all $(n+1)$-tuples of complex numbers (not all zero) with the equivalence relation

$$(z^0, \cdots, z^n) \sim (\lambda z^0, \cdots, \lambda z^n) \qquad \text{with} \qquad \lambda \in \mathbb{C} \setminus 0\,. \tag{5.77}$$

For the subset U_i of all equivalence classes in which $z^i \neq 0$, a chart is provided by

$$\phi_i : \{\,\text{class of } (z^0, \cdots, z^n)\,\} \;\mapsto\; \left(\frac{z^0}{z^i}, \cdots, \frac{z^{i-1}}{z^i}, \frac{z^{i+1}}{z^i}, \cdots, \frac{z^n}{z^i}\right) \in \mathbb{C}^n\,. \tag{5.78}$$

It is easy to show that these charts form an atlas and give explicitly the (holomorphic) transition maps. A Kahler potential in U_i is provided by

$$K^{(i)}(x) = \frac{1}{2}\,\ln\left(1 + \sum_{j=1}^{n} |x^j|^2\right)\,, \qquad \text{with} \qquad \{x^1, \cdots, x^n\} = \left\{\frac{z^0}{z^i}, \cdots, \frac{z^{i-1}}{z^i}, \frac{z^{i+1}}{z^i}, \cdots, \frac{z^n}{z^i}\right\} \tag{5.79}$$

the coordinates defined above. A straightforward calculation shows that this gives rise to a globally defined Kahler form and metric, the **Fubini-Study metric**. To be very concrete, it is easy to check that $\mathbb{C}P^1$ is the Riemann sphere. Crucially, all $\mathbb{C}P^n$ are compact.

Generally, submanifolds can easily be given as zero sets of polynomials, such as $x^2 + y^2 - 1$ on $\mathbb{R}^2$. The naive generalization to (holomorphic) polynomials on $\mathbb{C}^n$ is not useful for us since the resulting submanifolds are always non-compact (for $n > 1$). This is due to a generalization of the maximum modulus theorem for analytic functions. However, starting from the compact space $\mathbb{C}P^n$, compact submanifolds *can* be defined by polynomials. For the zero set to be well defined on the set of equivalence classes, the polynomials have to be homogeneous. Now, it can be shown that the crucial Calabi-Yau condition, the vanishing of the 1st Chern class, depends on the homogeneity degree of the polynomial. If we want to get a 3-fold, we must start from $\mathbb{C}P^4$. The Chern class vanishes if an only if the defining polynomial is of degree 5:

$$P_5(z) = c_{i_1 \cdots i_5} z^{i_1} \cdots z^{i_5}\,, \tag{5.80}$$

with indices running from 0 to 4 and labelling the projective coordinates of $\mathbb{C}P^4$. The so called **quintic** is then defined as the zero set,

$$P_5(z) = 0\,, \qquad z \in \mathbb{C}P^4\,. \tag{5.81}$$

Here by z we mean both the set of 5 numbers $\{z^i\}$ and the corresponding point in the projective space. As the coefficients of the polynomial vary, the complex structure changes. A concrete example is given, e.g., by

$$P_5(z) = (z^0)^5 + \cdots + (z^4)^5\,. \tag{5.82}$$

It is interesting to count the possible deformations such a quintic hypersurface: One first notes that the number of different monomials in a homogeneous polynomial of degree d in n variables is given by the binomial coefficient[11]

$$\begin{pmatrix} d+n-1 \\ n-1 \end{pmatrix} . \tag{5.83}$$

In our case this gives

$$\begin{pmatrix} 5+5-1 \\ 5-1 \end{pmatrix} = \begin{pmatrix} 9 \\ 4 \end{pmatrix} = 126 \, . \tag{5.84}$$

From this, we have to subtract the 25 parameters of the symmetry group $GL(5,\mathbb{C})$ of $\mathbb{C}P^4$, giving us 101 parameters. Recalling what was said before about the interplay of Dolbeault coholomogy and complex structure moduli spaces, we conclude that $h^{2,1}(Quintic) = 101$. Without derivation we also note that the Kahler form of $\mathbb{C}P^4$ is unique up to scaling, such that $h^{1,1} = 1$. Thus, for the quintic the Hodge diamond reads

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & 1 & & 0 & \\ 1 & & 101 & & 101 & & 1 \\ & 0 & & 1 & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array} \tag{5.85}$$

and the real dimension of the moduli space is $2 \cdot 101 + 1 = 203$.

We note that the same construction goes through for the quartic polynomial in $\mathbb{C}P^3$, giving rise to the unique Calabi-Yau 2-fold, known as the K3-surface. However, for 3-folds there are many more examples. First, one can generalize to the intersection of hypersurfaces (defined by polynomials) in products of projective spaces (giving rise to so-called complete-intersection CYs or CICYs). Then one can generalize from projective space to weighted projective spaces, in which case the different variables scale differently with a complex parameter λ. Furthermore, one may mod out not just by the rescaling by one such complex parameter, but by several such scalings (with different paramters λ_i). This leads to the concept of toric geometry and toric hypersurfaces, in which Calabi-Yaus can again be defined by polynomials of suitable degrees in the different variables (Batyrev's construction). Even more general Calabi-Yau constructions exist. The total number of distinct examples is about half a billion: $\sim 5 \times 10^8$.

5.6 Explicit parameterization of Calabi-Yau moduli spaces

We start with an extremely simple toy model: T^2. We can give it a complex structure by defining it as $\mathbb{C}/\mathbb{Z}^2$. By this we mean starting from the complex plane and modding out a lattice of translations, generated by unity and $\tau \in \mathbb{C}$. The resulting set of independent points, the so-called fundamental domain, is shown in Fig. 21. It is parameterized, on the one hand, by z and, on the other hand, by $x, y \in [0, 1)$, with the relation

$$z = x + \tau y \, . \tag{5.86}$$

[11] Try to prove this!

The complex number τ determines the complex structure. Note that tori with different τ are (in general) not isomorphic as complex manifolds. The holomorphic $(1,0)$-form in this case is clearly

$$\Omega = \alpha\, dz = \alpha\, dx + \alpha\, \tau\, dy\,, \tag{5.87}$$

with $\alpha \in \mathbb{C}$ an arbitrary constant.

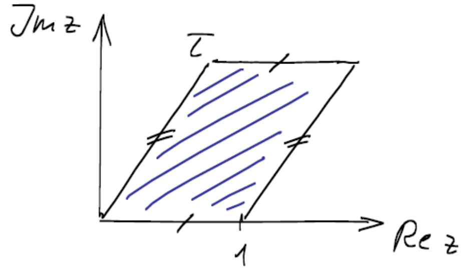

Figure 21: Torus defined as $\mathbb{C}/\mathbb{Z}^2$.

Now, in analogy to the proper Calabi-Yau case, the complex structure can be defined using the position of Ω in the complexification of $H^1(T^2)$. For this, it is sufficient to know the **periods**, i.e. the integrals of Ω over the integral 1-cycles:

$$\Pi_1 = \int_{y=\text{const.}} \Omega = \int_0^1 \alpha\, dx = \alpha \;\;, \qquad \Pi_2 = \int_{x=\text{const.}} \Omega = \int_0^1 \alpha\, \tau\, dy = \alpha\tau\,. \tag{5.88}$$

They can be combined in the period vector $\Pi = (\Pi_1, \Pi_2)$. Since the normalization of Ω is arbitrary, only ratios of these periods are meaningful. Concretely, the (in this single) complex structure parameter is given by $\tau = \Pi_2/\Pi_1$.

Next, we come to the moduli (in this case the modulus) associated with the Kahler form. The Kahler form is harmonic and can be decomposed in a basis of harmonic 2-forms,

$$J = t^i \omega_i\,. \tag{5.89}$$

Here the ω_i are in general chosen to represent an integral 2-form basis (where by integral we mean Poincare dual to the naturally defined integral basis of 4-cycles or, what is the same, the dual basis to the integral 2-cycle basis). In our case there is of course only one such 2-form:

$$\omega_1 = dx \wedge dy \;\;, \qquad \text{such that} \qquad J = t\, dx \wedge dy\,. \tag{5.90}$$

At the same time, we know that

$$J = i g_{i\bar{\jmath}}\, dz^i \wedge d\bar{z}^{\bar{\jmath}} = i g_{z\bar{z}}\, dz \wedge d\bar{z} = i g_{z\bar{z}}(dx \wedge \bar{\tau}\, dy - \tau\, dy \wedge dx) = -i(\tau - \bar{\tau})\, g_{z\bar{z}}\, dx \wedge dy\,. \tag{5.91}$$

Hence, we identify t as $t = -i(\tau - \bar{\tau})\, g_{z\bar{z}}$ and, recalling the general metric form

$$ds^2 = 2 g_{z\bar{z}}\, dz\, d\bar{z} = 2 g_{z\bar{z}}\big[dx^2 + |\tau|^2 dy^2 + (\tau + \bar{\tau})\, dx\, dy\big]\,, \tag{5.92}$$

we can finally explicitly give the matrix form of the metric in terms of the parameters t, Π_1, Π_2 which govern the position of J and Ω in their respective cohomology groups:

$$g_{ab} = 2g_{z\bar{z}} \begin{pmatrix} 1 & \mathrm{Re}\,\tau \\ \mathrm{Re}\,\tau & |\tau|^2 \end{pmatrix} = \frac{t}{\mathrm{Im}(\Pi_2/\Pi_1)} \begin{pmatrix} 1 & \mathrm{Re}(\Pi_2/\Pi_1) \\ \mathrm{Re}(\Pi_2/\Pi_1) & |\Pi_2/\Pi_1|^2 \end{pmatrix}. \tag{5.93}$$

With somewhat more writing, one can achieve the same level of explicitness for the toy-model 3-fold $T^2 \times T^2 \times T^2$, defined by modding out an appropriate lattice of translations from $\mathbb{C}^3$. Nevertheless, this is not a proper Calabi-Yau since its holonomy group is trivial. By contrast, a Calabi-Yau should have holonomy group $SU(3)$ (not just a subgroup). However, this is clearly to some extent a matter of convention. More importantly, T^6 is too simple for most physical applications and it does not give rise to the large landscape of solutions of string theory that we are after.

Thus, we now turn to the general case of proper Calabi-Yau 3-folds, such as the quintic and similar, even more complicated examples. The complete explicitness of metric parameterization that we saw above can of course not be achieved in such cases. But our main goal for the moment will be a description in 4d supergravity language,

$$\mathcal{L} = K_{i\bar{\jmath}}(\partial X^i)(\partial \overline{X}^{\jmath}) + \text{ gauge, fermion, and other fields}\,, \tag{5.94}$$

where K is the Calabi-Yau metric on moduli space, parameterized by the X^i, which include both Kahler and complex structure moduli. This *can* be given rather explicitly, even in the proper Calabi-Yau case.

Let us start with the Kahler moduli. As we already explained,

$$J = t^\alpha \omega_\alpha \qquad \text{with} \qquad \alpha = 1, \cdots, h^{1,1}\,. \tag{5.95}$$

Moreover, the volume of the Calabi-Yau can be given as

$$\mathcal{V} = \frac{1}{6}\int_X J \wedge J \wedge J = \frac{1}{6}\kappa_{\alpha\beta\gamma}t^\alpha t^\beta t^\gamma\,. \tag{5.96}$$

Here one may intuitively think of components the vector t^α as measuring the volumes of the different 2-cycles present in the Calabi-Yau. The integers $\kappa_{\alpha\beta\gamma}$ are the so-called triple intersection numbers of the 4-cycles Poincare dual to the ω_α.[12] The volumes of the dual 4-cycles, which are also labelled by the index α, are given by

$$\tau_\alpha = \frac{1}{2}\int_{c_4^\alpha} J \wedge J = \frac{1}{2}\kappa_{\alpha\beta\gamma}t^\beta t^\gamma\,. \tag{5.97}$$

Clearly, the variables t^α and τ_α encode the same information. Using them as $\mathcal{N} = 1$ SUGRA variables corresponds to choosing either of two different $\mathcal{N} = 1$ sub-algebras of the $\mathcal{N} = 2$ SUSY of a Calabi-Yau compactification of type IIA or IIB string theory. We focus (for the purpose of

[12]Note that 2 4-cylces in a 6d manifold intersect in a 2d submanifold. The latter intersects the 3rd 4-cycle in points. The total number of those, with orientation, is a function of the homology classes and is counted by the κs.

our later discussion of a particularly well-understood model, called KKLT) on the IIB case and the τ variables. They are real but, in 4d SUSY, are complexified by adding the imaginary parts

$$c_\alpha = \int_{c^\alpha} C_4\,. \tag{5.98}$$

Only as a side remark we note that, in the other SUSY, the t^α would be complexified by corresponding integrals of B_2 or C_2, depending on the particular model (for many more details see e.g. [75–78]).

Thus, we can solve for the t^α:

$$t^\alpha = t^\alpha(\tau_1, \cdots, \tau_{h^{1,1}}) \tag{5.99}$$

and, with $T_\alpha = \tau_\alpha + ic_\alpha$ and

$$\tau_\alpha = \frac{1}{2}(T_\alpha + \overline{T}_{\overline{\alpha}})\,, \tag{5.100}$$

we are finally able to give the type-IIB Kahler moduli Kahler potential:

$$K_K = -2\ln\mathcal{V} \qquad \text{with} \qquad \mathcal{V} = \mathcal{V}(T_\alpha, \overline{T}_{\overline{\alpha}})\,. \tag{5.101}$$

To describe the complex structure moduli space, we start by recalling the basis of $H_1(R_2)$ as given in Fig. 19. We rename the relevant cycles (representatives of the corresponding cohomology classes) as

$$a \to A^1\,, \qquad b \to B_1\,, \qquad c \to A^2\,, \qquad d \to B_2\,. \tag{5.102}$$

It is clear that this carries over analogously to the 1-cycles of higher Riemann surfaces, giving rise to the a basis $\{A^a, B_a\}$ and an intersection structure

$$A^a \cdot A^b = 0\,, \qquad B_a \cdot B_b = 0\,, \qquad A^a \cdot B_b = \delta^a{}_b\,. \tag{5.103}$$

An analogous basis can be chosen for the (in this case naturally isomorphic) vector space H_1. Such bases are called **symplectic bases**, on account of the antisymmetry of the only non-vanishing intersection numbers or, on the form side, wedge products:

$$\int \omega_a^A \wedge \omega_B^b = \delta_a{}^b\,. \tag{5.104}$$

The crucial point for us is that this represents a generic feature of the so-called **middle homology** or **cohomology** for manifolds where the dimensionalities of the relevant cycles/forms are odd. This is true for Riemann surfaces, with which we started, but it is equally true for complex 3-folds, our new case of interest.

Thus, now in the context of Calabi-Yaus, we choose a symplectic 3-cycle basis as above and define the periods

$$z^a = \int_{A^a} \Omega\,, \qquad \mathcal{G}_b = \int_{B_b} \Omega\,. \tag{5.105}$$

The complex paramters z^a with $a = 0, \cdots, h^{2,1}$ are sufficient to fully parameterize the position of Ω in $H^3(X)$. In fact, one of the parameters can be set to unity at the expense of a constant, complex rescaling of Ω, which does not induce any physical (geometrical) change. Hence one may think all the zs together as of 'projective coordinates'. Alternatively, one can set $z_0 = 1$, with $h^{2,1}$ parameters remaining.

Crucially, the remaining periods $\mathcal{G}_b$ are not independent - they are in general complicated functions of the zs:

$$\mathcal{G}_b = \mathcal{G}_b(z^0, \cdots, z^{h^{2,1}})\,. \tag{5.106}$$

One combines all of them in the period vector

$$\Pi = (z^0, \cdots, z^{h^{2,1}}, \mathcal{G}_0(z), \cdots, \mathcal{G}_{h^{2,1}}(z))\,. \tag{5.107}$$

The explicit form of the (dependent) periods can be obtained from appropriate differential equations (the **Picard-Fuchs equations**) which can be formulated on the basis of certain topological features of the Calabi-Yau. Crucially, they do not require the in general unavailable metric information. Thus, though with much work, the periods can in principle be explicitly obtained.

With this, we are ready to give the complex structure Kahler potential:

$$K_{cs} = -\ln(i\int_X \Omega\wedge\overline{\Omega}) = -\ln(-i\Pi^\dagger\Sigma\Pi) = -\ln(-i\overline{z}^a\mathcal{G}_a(z) + iz^a\overline{\mathcal{G}_a(z)})\,, \tag{5.108}$$

where

$$\Sigma = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \tag{5.109}$$

is the symplectic metric. See e.g. [79] for a nice summary and explanation of these and other, related formulae.

Finally, one non-geometric modulus related to the dilaton is generally present. It is known as the axio-dilaton (on account of the periodic scalar C_0):

$$S = C_0 + ie^{-\phi} = C_0 + \frac{i}{g_s}\,. \tag{5.110}$$

With this, the full type-IIB moduli Kahler potential (corresponding to a so-called orientifold projection with $O3/O7$ planes - the projection to $\mathcal{N} = 1$ mentioned earlier) reads

$$K = K_K(T^\alpha, \overline{T}^{\overline{\alpha}}) + K_{cs}(z^a, \overline{z}^{\overline{a}}) - \ln(-i(S-\overline{S}))\,. \tag{5.111}$$

This defines a 'ready-to-use' 4d supergravity model, so far without any scalar potential. The conventions are such that $M_{P,4} = 1$, as usual in supergravity lagrangians, and that fields measuring distances or volume in compact space (in our case the Ts) are doing so in string units, i.e. powers of $l_s \equiv 2\pi\sqrt{\alpha'}$. In other words, the above is valid with $l_s = 1$ concerning the Calabi-Yau geometry and $M_P = 1$ concerning 4d physics.

6 The flux landscape

The general idea will be to consider compactifications with non-zero internal components of the RR and NS field strength tensors F_p and H_3. This induces a non-zero superpotential depending on the moduli of the supergravity models discussed above and leads to moduli stabilization. Moreover, the number of available distinct models jumps from 10^8 to 10^{500} (or, in more general geometries – roughly speaking including D-branes – even much higher).

6.1 Compact geometries with p-form fluxes

Let us start with a few general comments on p-form gauge theories. Consider a (p–1)-form gauge theory in d dimensions, with an action of type (we disregard purely numerical prefactors)

$$\int \frac{1}{g^2}\, F_p \wedge *F_p + \int_{(p-2)-brane} A_{p-1}\,. \tag{6.1}$$

One can easily show that a **dual** description is provided by a theory based on the $(d-p)$-form field strength $\tilde{F}_{d-p}$. The latter is defined as

$$\tilde{F}_{d-p} = \frac{1}{g^2} * F_p\,, \tag{6.2}$$

which in turn leads to the definition of a dual gauge potential via

$$\tilde{F}_{d-p} = d\tilde{A}_{d-p-1}\,. \tag{6.3}$$

In these new variables the action takes the form

$$\int \frac{1}{\tilde{g}^2}\, \tilde{F}_{d-p} \wedge *\tilde{F}_{d-p} + \int_{(d-p-2)-brane} \tilde{A}_{d-p-1} \qquad \text{with} \qquad \tilde{g} = \frac{1}{g}\,. \tag{6.4}$$

While the new kinetic term is just a rewriting of the old one, the charged objects coupling to the dual potential are different. In fact, both types of charged objects are present in the full theory. But the coupling of any one of them to the fields can only be explicitly given on one side of the duality.[13]

The above is of course familiar from electrodynamics, where $d = 4$ and $p = 4 - p = 2$, such that the tilde is really necessary to distinguish the otherwise identical-looking dual descriptions. The charged objects on both sides are 0-branes, i.e. particles.

Now let us consider the particularly simple case of F_1 in $d = 4$, which is of course nothing but a scalar (axion) field model, with $A_0 \equiv \phi$:

$$\int f^2(\partial\phi)^2 \;+\; \phi(x_i)\,. \tag{6.5}$$

The last term is the coupling to an instanton, a tunneling event localized at a point at a point in spacetime. Clearly, such objects can not be included in an initial field configuration – one has to sum over them and integrate over all the x_i in the path integral.

In case this concept is unfamiliar, here is a brief excursion concerning **instantons**. The 'classical' setting is in fact that of a gauge theory coupled to a periodic pseudo-scalar or axion-like-field or axion for short. For definiteness, say the gauge group is $SU(2)$:

$$\mathcal{L} = \frac{1}{2g^2}\,\mathrm{tr}\, F_{\mu\nu}F^{\mu\nu} + \frac{\phi}{8\pi^2}\,\mathrm{tr}F \wedge F\,. \tag{6.6}$$

[13] The purely classical dualization above can also be performed at the quantum level, i.e. under the path integral. The basic idea is to implement the original Bianchi idenity constraint $dF_p = 0$ by a lagrange multiplier, i.e. by adding a term $dF_P \wedge A_{d-p-1}$ to the action. Then one can integrate out F_p, arriving at the dual action. In the latter, the lagrange multiplier A_{d-p-1} has become the new dynamical field.

The term multiplying ϕ is a total derivative but there exist field configurations (which can not be smoothly deformed to the vacuum) on which the integral gives $8\pi^2 n$ with $n \in \mathbb{Z}$. Very roughly speaking, the existence of such a field configuration is related to the fact that $SU(2) \cong S^3$ and the possibility of identifying this group-theoretic S^3 with the S^3 of radial coordinates in $\mathbb{R}^4$. In the euclidean path integral, one has to sum over all such 'bumps of energy-density' (to be interpreted as local tunneling events, leading from vacuum to vacuum). One also has to integrate over all their sizes $1/M$ and positions. The events are suppressed by their action – $\exp(-S_i)$ – and for large S_i one uses the 'dilute gas approximation' (cf. Fig. 22). It should now be clear in which sense our model of (6.5) corrsponds to instantons of an $SU(2)$ (more generally $SU(N)$) gauge theory: The point at which the gauge-field-theoretic instanton is localized is identified with x_i and the $F \wedge F$ term of the lump of field strength is replaced by an approximate δ-function.

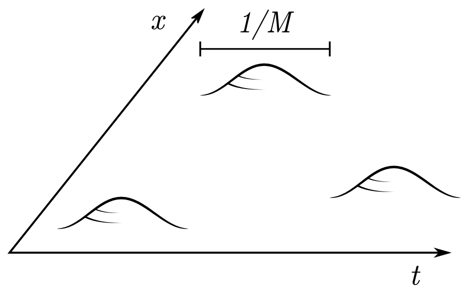

Figure 22: Instantons as localized lumps of field strength (figure from [80]).

Still within our excursion about instantons, we recall that a model with a periodic scalar like that of (6.5) can be derived by compactifying a 5d $U(1)$ gauge theory to 4d. Interestingly, this also has instantons, but of a very different type (cf. Fig. 23). We leave it as an excercise for the reader to derive the correct coupling of this type of instanton to ϕ. This ends our instanton excursion.

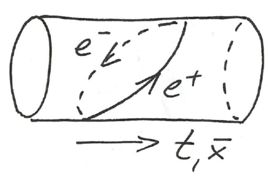

Figure 23: Effective instanton arising from a particle-antiparticle fluctuation wrapping the compact space of an S^1 compactification (figure from [80]).

As a side remark, the dual theory has field strength $H_3 = dB_2$ and couples to strings (which are here unrelated to any fundamental string theory). What interests us here is flux quantization, which is particularly easy to understand in this case. Indeed, if our spacetime has non-trivial one-cycles, the gauge potential ϕ does not need to be globally well defined but it can (assuming e.g. that x_3 paramerizes an S^1) obey

$$\phi(x_i) = \phi(x_i + 2\pi R) + 2\pi n\,, \qquad n \in \mathbb{Z}\,. \tag{6.7}$$

The shift must be integer or else the instanton action would not be well-defined. Another way to formulate the same condition is

$$\oint F_1 = 2\pi n\,. \tag{6.8}$$

Here n is either the number of fundamental strings wrapped by the loop or it is a discrete choice one has to make when defining the theory on a spacetime with a fundamental 1-cycle.

The above is clearly analogous to the familiar statement

$$\oint_{S^2} F_2 = 2\pi n \tag{6.9}$$

for electrodynamics and an S^2 enclosing n magnetic monopoles. But this case is not our interest at present. What we care about is flux quantization,

$$\oint_{c_p} F_p \in 2\pi\,\mathbb{Z}\,, \tag{6.10}$$

which is simply a requirement of (quantum-mechanical) consistency of a p-form gauge theory (and its dual). In the absence of charges, the flux can only be non-zero if a non-trivial p cycle exists in the geometry.

Now let us compactify a 4d model with a 0-form gauge theory (an axion) to 3d on S^1. The compact geometry has a single compact 1-cycle. This allows a choice of boundary conditions or, equivalently, 1-form fluxes on the S^1. The freedom is precisely that of choosing $n \in \mathbb{Z}$ in (6.7). Thus, one obtains an infinity of 3d models with different vacuum energy: Indeed, from the 3d perspective, the gradient term $(\partial_3\phi)^2$ contributes to the cosmological constant. This already represents a small flux landscape. Moreover, the theory possesses strings. Let us include an infinite string in our compactification. This string is a point in the compact x^3-direction and hence still has two dimensions - one time and one spacelike, in the noncompact (2+1)-dimensional spacetime. It is hence a domain-wall in the noncompact 2d space. Once can convince oneself that, on the two sides of this wall, the flux on the S^1 differs by one unit. Hence our flux landscape is actually not just a collection of different theories, but it possesses a dynamics allowing one to change between those: This dynamics is **bubble nucleation** (cf. Fig. 24). The surfaces of the bubbles are the domain walls made of the higher-dimensional charged objects. This crucial feature will survive in the full-fledged string theory landscape.

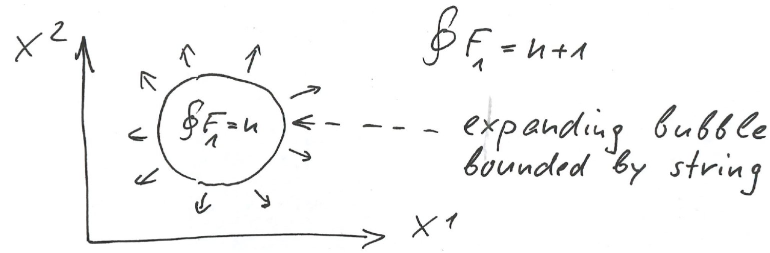

Figure 24: Bubble nucleation in a 4d-to-3d toy model with 1-form flux.

Clearly, an analogous situation may be considered if one compactifies, for example, a 6d gauge theory to 4d on S^2. The S^2 may be given 2-form flux in the sense of (6.9), giving rise to a 4d landscape of vacua labelled by n. In this case, the flux quantization is literally based on the same logic that forces the F_2 integral around a magnetic monopole to be quantized. One may also use the $U(1)$ principle bundle approach to gauge theories to think of this in terms of non-trvial fibrations of $U(1)$ over S^2, which are known to be labelled by an integer, our flux number. The case of unit flux corresponds to the famous Hopf fibration (see e.g. [81, 82]).

6.2 Bousso-Polchinski model

Now let us look for an effective field theory [83] for this apparently rather general mechanism of creating 'landscapes' directly in 4d. Such an effective description arises naturally if consider the (somewhat special) case of a $(d-1)$-form gauge theory in d dimensions. Let us for concreteness focus on a 3-form gauge theory in $d = 4$:

$$S \sim -\int \frac{1}{\Lambda^4} F_4^2 + \int_{\text{domain wall}} A_3 \,. \tag{6.11}$$

Without sources, the equation of motion $d * F_4 = 0$ implies that F_4 is constant, so there are no propagating degrees of freedom. The only dynamics is that of domain walls, which have some tension and hence move according to their own classical dynamics. Moreover, they couple to A_3 and hence source F_4.

With the domain wall comes a 1-form current, appearing in

$$\frac{1}{\Lambda^4} d * F_4 = j_1 \,, \tag{6.12}$$

which is localized at the wall. As is generally the case in p-form gauge theories, the integral of the current counts the number of charged objects. In the most familar case of 4d electrodynamics, the integral of the 3-form current over a spatial 3d volume counts the number of charged-particle worldlines crossing that volume. Here, our 1-form current should integrate to unity on any line that crosses the domain wall once. Concretely, consider a finite line, with beginning and end point on opposite sides of the wall, such that

$$1 = \int_{Line} j_1 = \frac{1}{\Lambda^4} \int_{Line} d * F_4 = \frac{1}{\Lambda^4} (*F_4) \Big|_{x_1}^{x_2} . \tag{6.13}$$

From this, we see right away that the scalar $*F_4$ jumps by Λ^4 when crossing the wall. The dual description, though even more exotic, is simpler:

$$S \sim -\int \Lambda^4 \, F_0^2 \,, \tag{6.14}$$

without any meaningful 'A_{-1}' or sources. The 0-form field strength is classically identified with $*F_4$, it is constant in spacetime by its Bianchi identity, $dF_0 = 0$, and it only takes discrete values. This follows from the solution for F_4 in the vicinity of a domain wall discussed above. It can also be viewed as a degenerate version of flux quantization. The set of vacua following from the F_0 description is displayed in Fig. 25.

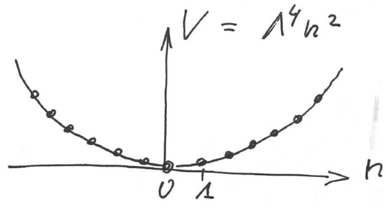

Figure 25: Flux discretuum of a 3-form gauge theory.

Now let us assume that our 4d theory possesses a large number of such 4-form fields,

$$S \sim - \int \sum_{i=1}^{N} \Lambda_i^4 \, F_{i,0}^2 \, . \tag{6.15}$$

This can arise, for example, if it originates from a compactification of a higher-dimensional p-form gauge theory on a compact space with N $(p+1)$-cycles. The flux on each of those cycles then corresponds to the flux number n in one of the F_0-models in (6.15). The two-field case is illustrated in Fig. 26.

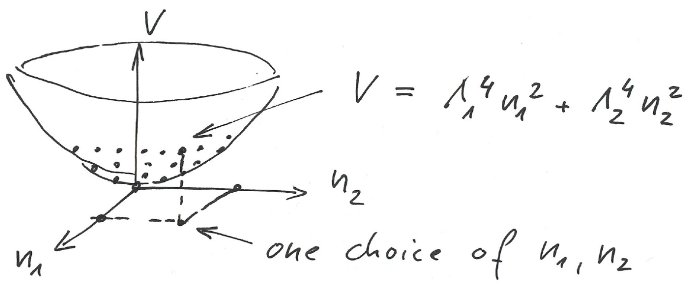

Figure 26: Flux discretuum of a 3-form gauge theory with two fields.

Each flux choice gives rise to a particular cosmological constant

$$\lambda(\overline{n}) \equiv V(\overline{n}) = \sum_i \Lambda_i^4 n_i^2 \, . \tag{6.16}$$

One may ask how many different flux choices lead to $\lambda(\overline{n}) < \lambda_0$. To simplify the discussion, let us assume that all 4-form gauge couplings are equal: $\Lambda_i = \Lambda$. The number of flux choices is then simply the number of lattice points $\overline{n} \in \mathbb{Z}^N$ inside a ball of radius $\sqrt{\lambda_0}/\Lambda^2$. The lattice is N-dimensional, so the desired number is

$$K(\lambda_0) \sim (\sqrt{\lambda_0}/\Lambda^2)^N \, . \tag{6.17}$$

If $\sqrt{\lambda_0} > \Lambda^2$, this grows exponentially fast with N. In particular, the number of points leading to

$$\lambda \in [\lambda_0, \lambda_0 + \delta\lambda] \tag{6.18}$$

will, for moderately large N (say $N = \mathcal{O}(100)$, as suggested by the number of 3-cycles of the quintic), still be extremely large. This remains true even if $\delta\lambda$ is very small:

$$\delta K(\lambda_0, \delta\lambda) \sim (\sqrt{\lambda_0}/\Lambda^2)^{N-1}(\delta\lambda/\sqrt{\lambda_0}\Lambda^2)\,. \tag{6.19}$$

Note that we do not have to be afraid that regularities in the distribution of λ-values could lead to intervals into which λ never falls: Such possible regularities will be destroyed if we make all Λ_i different, as expected in a more realistic situation.

So far, we have a model with many solutions. These solutions give rise to a discretuum a cosmological constants, which becomes extremely dense in the region $\lambda \gtrsim \Lambda^4$ (where Λ sets the typical scale for the couplings Λ_i). Now, by adding a negative cosmological constant $\lambda_{AdS} < 0$, such that

$$S \sim -\left(\int \sum_{i=1}^{N} \Lambda_i^4\, F_{i,0}^2 \;+\; \lambda_{AdS}\right)\,, \tag{6.20}$$

we can shift this dense discretuum downward. In this model, we are statistically guaranteed that vacua with an extremely small cosmological constant exists. Clearly, due to the possible bubble nucleation processes these vacuum will only be metastable, but they can be very long-lived. We will play with numbers later on to see *how* small $\lambda(\overline{n})$ can really become.

6.3 The type-IIB flux landscape (GKP)

The key idea or observation is that, in type IIB Calabi-Yau compactifications, the 3-form fluxes of H_3 and F_3 can play roughly the role of the multiple fluxes of the Bousso-Polchinski model discussed above. The details are, however, more complicated and in part qualitatively different, mainly due to the central role of supersymmetry.

We start with the intuitve observation that a non-zero flux on a compact cycle (say a 1-cycle) clearly has an energetic effect. Indeed, let us for simplicity assume that the compact space is $S^1_A \times S^1_B$ and one unit of 1-form flux sits on the B-cycle. Then

$$\int dy_B\, F_1 = 1 \qquad \text{and hence} \qquad F_1 \sim 1/R_B\,. \tag{6.21}$$

This gives rise to a contribution to the action

$$S \supset -\int d^4x \int dy_A\, dy_B\, F_1 \wedge *F_1 \sim -\int d^4x\,(R_A R_B)\cdot\frac{1}{R_B^2} \sim -\int d^4x\,\frac{R_A}{R_B}\,. \tag{6.22}$$

We learn that a flux on a cycle prevents this cycle from shrinking. More generally, if there are fluxes of various values on various cycles of a compact space, then these fluxes tend to stabilize the shape of the manifold in a certain way. Specifically, the ratio between the volumes of two cycles gets stabilized roughly according to the ratio of the flux numbers on these cycles.

Concretely, we expect that 3-form fluxes will stabilize (give mass to) the complex structure moduli, which a we know govern the ratios of 3-cycle volumes. But this is not possible in a 4d supergravity model without superpotential since for $W = 0$ no scalar potential is induced.

To make the right guess for the form of the expected flux-induced W, it is useful to observe that (already in 10d) one can use the complex scalar field

$$S = C_0 + ie^{-\phi} \tag{6.23}$$

and the complex 3-form flux

$$G_3 = F_3 - S\,H_3\,. \tag{6.24}$$

The kinetic terms of the two 3-form fields take the simpler form (suppressing constant prefactors)

$$S \supset \int d^{10}x\, G_3 \wedge *\overline{G}_3\,. \tag{6.25}$$

With this, one may guess the mathematically natural expression for the superpotential induced by 3-form fluxes:

$$W = \int_X G_3 \wedge \Omega_3\,. \tag{6.26}$$

This is known as (the type IIB version of) the Gukov-Vafa-Witten superpotential [84]. The latter has first been postulated and mathematically justified (in an abstract way) for M-theory compactifications to 3d on Calabi-Yau 4-folds:

$$W_{GVW} = \int_{X_4} G_4 \wedge \Omega_4\,. \tag{6.27}$$

In the famous paper for Giddings, Kachru and Polchinski (GKP) [63] (see also [85]), this superpotential was used and justified explicitly by comparing the 4d scalar potentials derived in from 4d $\mathcal{N}=1$ supergravity and directly from 10d.

Now one can make this fully explicit by normalizing the 3-form fields such that flux quantization takes the form

$$\frac{1}{2\pi\alpha'}\int F_3 \in 2\pi\mathbb{Z}\,, \qquad \frac{1}{2\pi\alpha'}\int H_3 \in 2\pi\mathbb{Z} \tag{6.28}$$

for integrals over integer cycles. Equivalently, one may decompose the fluxes in a symplectic integer form basis,

$$F_3 = -(2\pi)^2\alpha'(f^a\omega_a^A + f_{b_3/2+b}\omega_B^b)\,, \qquad H_3 = -(2\pi)^2\alpha'(h^a\omega_a^A + h_{b_3/2+b}\omega_B^b)\,, \tag{6.29}$$

where the entries of the coefficients vectors f and h now have to be integer. With this the superpotential, given in its simplest and mathematically natural form above, can be worked out explicitly:

$$W = \int_X G_3 \wedge \Omega_3 = (2\pi)^2\alpha'\,(f - Sh)\cdot\Pi(z)\,. \tag{6.30}$$

The scalar potential potential reads, as usual,

$$V = e^K\left(K^{i\bar{\jmath}}(D_iW)(D_{\bar{\jmath}}\overline{W}) + K^{\alpha\overline{\beta}}(D_\alpha W)(D_{\overline{\beta}}\overline{W}) - 3|W|^2\right)\,. \tag{6.31}$$

Here we have, for simplicity and since they all appear in $W = W(S,z)$, combined the axio-dilaton S and the complex structure moduli $z^1, \cdots, z^{h^{2,1}}$ in one vector:

$$z^i = \{S, z^1, \cdots, z^{h^{2,1}}\}\,. \tag{6.32}$$

We have furthermore redefined

$$K_{c.s} - \ln(-i(S+\overline{S})) \;\to\; K_{c.s}\,, \tag{6.33}$$

absorbing the axio-dilaton Kahler potential into the complex-structure Kahler potential.

Since W is independent of the Ts and since the Kahler modulus Kahler potential is of no-scale type, e.g. in the simplest case

$$K_K = -2\ln(\mathcal{V}) = -2\ln((T+\overline{T})^{3/2}) = -3\ln(T+\overline{T})\,, \tag{6.34}$$

the last two terms in (6.31) exactly cancel (see problems):

$$V = e^K\, K^{i\bar{\jmath}}(D_iW)(D_{\bar{\jmath}}\overline{W})\,. \tag{6.35}$$

Moreover, the equations for unbroken SUSY (the F-term conditions)

$$D_iW = 0 \qquad \text{for} \qquad i = 1, \cdots, b_3/2\,, \tag{6.36}$$

represent $b_3/2$ equations for equally many complex variables. They will in general possess a solution (or a finite set of solutions). This fixes all z^i to specific values. One may view these fields, which now have a large mass in this positive definite potential, as being integrated out. The result is a model depending just on T (or, more generally, all Kahler moduli) in which

$$V = V(T,\overline{T}) \equiv 0\,. \tag{6.37}$$

Since

$$\overline{F}^{\overline{T}} = D^{\overline{T}}W = K^{\overline{T}T}K_TW = \left(\frac{3}{(T+\overline{T})^2}\right)^{-1}\frac{(-3)}{(T+\overline{T})}\,W = -(T+\overline{T})\,W \neq 0\,, \tag{6.38}$$

supersymmetry is broken. The scale at which it is broken (e.g. the gravitino mass $e^{K/2}W$) is not fixed since T is not fixed. One calls this a **no-scale model**.

One of the key points of [63] (known as ‘GKP’) is that they established this vanishing potential not only (as we just did) indirectly, via 4d SUGRA arguments, but by explicitly providing the 10d geometry. The term ‘explicitly’ is here interpreted as follows: One assumes that a Calabi-Yau metric is given (this is of course not explicit but rests on the famous existence theorem). Then, given in addition certain fluxes and other sources in the Calabi-Yau (e.g. O3-planes and D3 branes), one is able to write down differential equations determining the actual metric, including backreaction from fluxes. This metric corresponds to a flux compactification to 4d Minkowski space. In fact, there is a family of such solutions, corresponding to the flat direction characterized by the ‘no-scale modulus’ T above.

6.4 Kahler modulus stabilization and SUSY breaking (KKLT)

As in the previous section, we focus on the simplest case $h^{1,1} = 1$, such that

$$K = -3\ln(T+\overline{T}) \qquad \text{and} \qquad W = W_0 = \text{const.} \tag{6.39}$$

The complex structure moduli have been integrated out and the corresponding flux choice (together with the VEVs of the z^i which it prescribes) has fixed W_0. At leading order, we have $V \equiv 0$ and SUSY breaking with $m_{3/2} \sim e^{K/2} W_0$.

Various (quantum) corrections will generically lift the flatness of V, breaking the no-scale structure. This can be α' corrections (corresponding to higher-dimension operators in 10d), loop-corrections, non-perturbative instanton effects or non-perturbative effects from (SUSY) gauge theory confinement (also known as gaugino condensation). The last two of these four qualitatively different effects lead to technically similar results. In particular, W is corrected according to

$$W_0 \qquad \to \qquad W_0 + A\, e^{-2\pi T/N}\,, \tag{6.40}$$

by either instantons (in this case $N = 1$) or by gaugino condensation (here N is the dimension of the fundamental representation of the gauge group $SU(N)$). This type of corrections is one of basic ingredients of the KKLT-scenario for (complete) moduli stabilization and SUSY breaking [86] which we will describe in the rest of this section.

We will not introduce the technology necessary to describe the case of gaugino condensation. Suffice it to say that, if a stack of N D7 branes is wrapped on a 4-cycle with volume $\sim \mathrm{Re}\, T$, the 4d theory contains a corresponding $\mathcal{N} = 1$ super-Yang-Mills theory. The latter exhibits confinement, as familiar from the non-SUSY QCD sector of the Standard Model. In our case, confinement is characterized by a non-zero $W_{\mathrm{non-pert.}} \sim \Lambda^3$. Here Λ is the confinement scale and its relation to $W_{\mathrm{non-pert.}}$ follows on dimensional grounds. Using also the fact that the 4d gauge coupling squared is $\sim 1/\mathrm{Re}\, T$ by the standard logic of Kaluza-Klein reduction from 8d to 4d, one may run from high to low energy scales and determine at which scale the gauge coupling reaches $\mathcal{O}(1)$ values. This fixes Λ in terms of $\mathrm{Re}\, T$, and by holomorphicity leads to $W_{\mathrm{non-pert.}} \sim \exp(-2\pi T/N)$, where N comes in through the beta-function.

Since have already introduced some of the ideas relevant for instanton effects, we will describe the instanton case in slightly more detail. For this, it is useful to recall how an instanton correction in a 5d-to-4d compactification of a gauge theory is related to the possibility of wrapping a closed electron worldline on a 1-cycle (in this case the unique 1-cycle S^1) of the compact space. This was illustrated in Fig. 23, where the reader was invited to think of a particular type of e^+e^- fluctuation of the vacuum. But this time-dependent picture is not necessary – the simplest and dominant effect corresponds to just wrapping the worldline on the minimal volume cycle (at fixed 4d space-time point x^μ), subsequently integrating over t.

This type of instanton has an obvious analogue in compactifications of higher-form gauge theories. The interesting case is that where the compact space possesses cycles the dimensions of which correspond to the dimensionsionality of the available charged objects. Our case of interest is type IIB with its 4-form C_4 and the corresponding D3 branes. We think of the electron worldline above as of a 0-brane and, once it is wrapped in euclidean signature to describe tunneling, as of an E0 brane. Analogously, we can think of a D3 brane (now often called an E3 brane) as being wrapped at fixed x^μ on the minimal volume 4-cycle of our Calabi-Yau. This is the origin of the instanton correction we are after, cf. Fig. 27. Such instantons are called stringy or exotic or D-brane instantons.

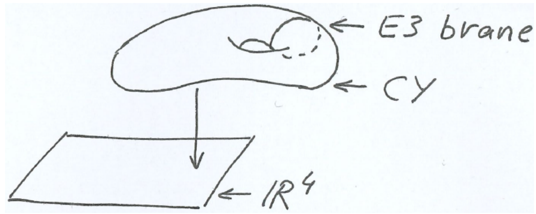

Figure 27: An E3-brane instanton, corresponding to a euclidean D3 brane wrapped on a 4-cycle of a CY over one of the points of the non-compact space-time $\mathbb{R}^4$.

At the quantitative level, we recall that our complex Kahler modulus is $T = \tau + ic$, where

$$\tau \sim R_{CY}^4 \sim \int_{4-cycle} \sqrt{g_{CY}}\,. \tag{6.41}$$

The last expression is, up to the proper normalization by the tension prefactor, the action of the wrapped brane. Furthermore, the wrapped brane couples to C_4 through

$$\int_{4-cycle} C_4 \sim c\,, \tag{6.42}$$

which is just the 4d axionic scalar in T. Thus, a single instanton contributes to the 4d partition function as

$$\sim e^{-2\pi\tau}\, e^{-2\pi i c}\,, \tag{6.43}$$

where the first factor is the tunneling suppression by the euclidean brane action. The second factor comes from the part of the D3-brane action displayed in the previous line. It can equivalently be viewed purely in 4d as the coupling of the 0-form gauge field c to its 0-dimensional charged object, the instanton.

Summing over all numbers of instantons and anti-instantons (which come with $e^{+2\pi i c}$) leads to an exponentiation:

$$\mathcal{L}_{4d} \;\supset\; \exp\left[\sim e^{-2\pi\tau}\,\cos(2\pi c)\right]\,. \tag{6.44}$$

The term in the exponent is the instanton correction to the 4d effective action and it is precisely analogous to the possibly more familiar gauge theory case. Here, one gets corrections $\sim e^{-8\pi/g^2}\cos(2\pi\phi)$, where g is for example the strong gauge coupling and ϕ the QCD axion, famously obtaining a cosine-potential from this effect.

In SUSY, such instanton corrections can enter the 4d effective action only through either K or W:

$$W_0 \quad\to\quad W_0 + Ae^{-2\pi T} \qquad\qquad \text{or} \qquad\qquad K \quad\to\quad K + Be^{-2\pi T} + \text{c.c.}\,. \tag{6.45}$$

Which of the two happens depends on the geometry of the wrapped brane and will not be discussed here. For KKLT, we require that a correction to W arises. We also note that the τ and c dependences are such that they combine in a holomorphic function of T (as required by SUSY), with the proper periodicity in $\mathrm{Im}(T)$. Conversely, as shown in the problems, the evaluation of the scalar potential on the basis of W from (6.45) leads to a term of the type of (6.44).

We can now finally proceed with the analysis of the 4d effective theory, defined by

$$K = -3\ln(T+\overline{T}) \qquad \text{and} \qquad W = W_0 + Ae^{-aT}\,. \tag{6.46}$$

It is a straightforward exercise to derive the scalar potential $V(\tau, c)$, integrate out c (by simply finding the minimum in c), and thus obtain

$$V = V(\tau)\,. \tag{6.47}$$

The qualitative behavior of this potential at $W_0 \ll 1$ is displayed in Fig. 28. It is easy to derive by analysing the standard supergravity formula for V in the regimes $e^{-aT} \gg W_0$ and $e^{-aT} \ll W_0$ (see problems). One checks that V grows at small τ and approaches zero from below at large τ. This is sufficient to conclude that the qualitative picture is that of Fig. 28.

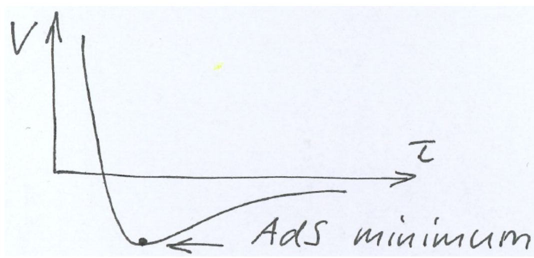

Figure 28: Qualitative behaviour of the scalar potential arising after the inclusion of instanton or gaugino condensation effects in W.

Moreover, is is easy to prove in general that the supergravity scalar potential has an extremum at supersymmetric points, where the F-terms and hence the first, positive-definite term in the supergravity potential formula vanish,

$$DW = -e^{-T} - \frac{3}{T+\overline{T}}\left(W_0 + e^{-T}\right) = 0\,. \tag{6.48}$$

In our case this extremum is always a minimum. Here we have set $A = a = 1$ for simplicity. Assuming $c = 0$, this vanishing-F-term condition is solved (implicitly in τ) if

$$W_0 = -\left(1 + \frac{2}{3}\tau\right)\, e^{-\tau} \tag{6.49}$$

holds. The conclusion that W_0 must be real and negative is a mere consequence of our simplifying assumption $c = 0$. For a general phase of W_0 (and A), we would simply have found a non-zero value of c at the minimum. This is not important for us.

What *is* important is the conclusion that W_0 must be exponentially small for parametric control, i.e. to have $R_{CY} \gg 1$. Of course, making W_0 small should not be a problem since it depends on the flux choice – it can hence be finely tuned in the landscape. In fact, to be sure that nothing goes wrong one needs to know that the statistical distribution of W_0 in the complex-W_0-plane for random flux choices has no special feature near the origin. This crucial fact, more precisely the flatness of the distribution of $|W_0|^2$ values near zero, has been established with some level of rigour in [87].

Thus, we have uncovered a landscape of supersymmetric vacua with a negative cosmological constant, so-called SUSY AdS vacua. (Note that, in the 'first step of KKLT' leading to these solutions the broken supersymmetry of GKP is restored in the minimum). But to describe the real world we need a positive (even though very tiny) cosmological constant and broken supersymmetry. Moreover, turning at least a small fraction of the SUSY-AdS vacua above into dS vacua is essential for eternal inflation, the presently leading mechanism for populating the landscape cosmologically (see below).

Let us first give a much simplified, 'macroscopic' description of how dS vacua may arise on the basis of the above. Let us assume that some further details of the model, such as branes with their gauge theories and charged matter fields, introduce extra light degrees of freedom and corresponding corrections to K and W:

$$K \quad \to \quad K(T\overline{T}) + \delta K(X, \overline{X}, T, \overline{T}) \quad , \qquad\qquad W \quad \to \quad W(T) + \delta W(X)\,. \tag{6.50}$$

Now, let us choose δK and δW in analogy to the one-fiel O'Raifeartaigh-type model discussed in Sect. 2.7:

$$\delta K \sim X\overline{X} - (X\overline{X})^2 \ , \qquad \delta W = \alpha X\,. \tag{6.51}$$

This will lead to $D_X W \neq 0$ in the vacuum. Moreover, one chooses parameters such that, in this SUSY breaking vacuum $X = 0$ and its fluctuations have a very large mass. Then the upshot of the whole construction is that the scalar potential V is supplemented by a so-called uplifting term

$$V \quad \to \quad V + \delta V \qquad\qquad \text{with} \qquad\qquad \delta V = e^K K^{X\overline{X}} |D_X W|^2\,. \tag{6.52}$$

At this generic level of analysis the uplifting term δV could have any T dependence, given our free choice of the T dependence of δK. In concrete string constructions, for which the above is a toy model, δV will always be decaying at large volume, cf. Fig. 29. This can be understood if one imagines that (as is mostly the case) δK and δW are due to some local effect in the CY. Then, going to large volume, the SUSY-breaking and uplifting effects stay the same in string units, but the Planck mass diverges. Hence, in standard supergravity conventions with $M_P = 1$, δV will decay with growing T.

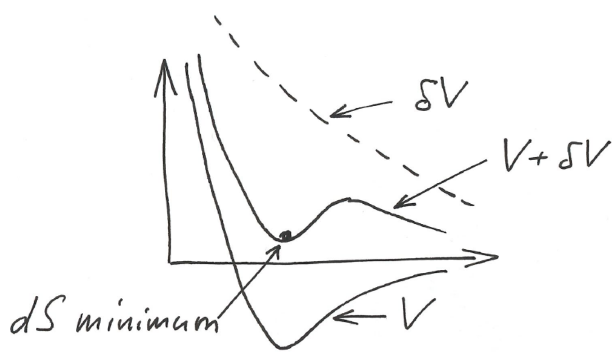

Figure 29: Uplifting to a KKLT dS vacuum.

One may expect that in the huge string theory landscape many options for such an uplift exist. Yet, it turns out not to be easy to construct an uplift of the above O'Raifeartaigh type explicitly. Thus, the most explicit uplift has a somewhat different structure: It is the anti-D3-brane uplift originally suggested by KKLT, which arguably remains the most explicit (though nevertheless not uncontroversial[14]) possibility. We turn to this construction, which requires some more technology, next. As we will see, even though different in detail, the KKLT uplift behaves qualitatively as explained using the O'Raifeartaigh toy model above.

6.5 The anti-D3-brane uplift of KKLT

As we explained earlier, a so-called orientifold projection reduces the supersymmetry of a type II CY compactification from $\mathcal{N} = 2$ to $\mathcal{N} = 1$. Let us consider the example of an O3-plane projection, which can locally be thought of as the geometric action

$$(z^1, z^2, z^3) \quad \to \quad (-z^2, -z^2, -z^3)\,, \tag{6.53}$$

to be combined with a world-sheet orientation change.

Locally, this projection introduces a singularity at $\{z^i\} = 0$, at which (due to the orientation change) a so-called O3-plane is localized. This is a negative-tension object which also has opposite C_4-charge[15] compared to a D3-brane. In a consistent compactification, an O3-plane always has to come with a certain number of D3 branes for total charge neutrality (tadpole cancellation). Concretely, the D3 charge of an O3-plane is -1/4. The fractionality is not a problem since the compact CY after orientifolding will usually have a large number, divisible by 4, of O3-planes. For example, it is easy to check that $T^6/\mathbb{Z}_2$, with the $\mathbb{Z}_2$ acting as above, has 64 O3-planes.

Now, given a consistent CY with a number of O3-planes and a corresponding number of D3-branes, it is possible to replace some or all of the D3-branes by 3-form fluxes. This possibility arises since, through the CS-term, 3-form fluxes contribute to the total D3 tadpole. This takes us to the realm of flux compactifications à la GKP and, if we also allow for the non-perturbative effects $\sim e^{-aT}$ introduced above, we will find ourselves in an $\mathcal{N} = 1$ SUSY setting with O3-planes, D3-branes and fluxes. The O3-planes, the D3-branes and the fluxes all break SUSY to the same $\mathcal{N} = 1$ sub-algebra of the original $\mathcal{N} = 2$ SUSY of the pure CY model.

Next, we can think of breaking SUSY by adding a D3 and anti-D3 (for short: $\overline{D3}$) brane pair. The $\overline{D3}$ breaks $\mathcal{N} = 2$ to the opposite $\mathcal{N} = 1$ subalgebra, such that 4d SUSY is now completely broken. D3 tadpole cancellation is not violated since we added two oppositely charged objects. However, brane and anti-brane attract each other both gravitationally and through C_4, so they will quickly find each other and annihilate, releasing twice the energy density of the D3-brane tension. Our 'uplift' is thus very short-lived and not practically useful.

However, we could avoid having any D3-branes by cancelling the tadpole of the O3-planes by flux alone. If we now add a $\overline{D3}$-brane and increase the flux appropriately to ensure tadpole cancellation, we appear to have the desired uplift. Now, the $\overline{D3}$ still breaks SUSY relative to flux and O3-planes but there is no D3 which it could attract and annihilate.

[14]There even exists the opinion that no uplift to a dS minimum can ever be constructed for fundamental reasons, challenging most ideas about how string theory might be relevant to the real world. We will return to this subject [88, 89].

[15]This is often referred to as D3-charge.

Unortunately, this is not yet good enough since this uplift (by twice the D3-brane tension, which is string-scale) is much too strong. Indeed, given that the non-perturbative effects and hence the depth of the original AdS minimum are exponentially small, the situation will be as in Fig. 30: The uplift is much too strong and no local dS minimum can be generated.

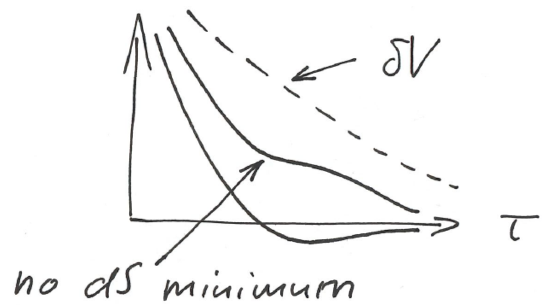

Figure 30: Too high an uplift.

Fortunately, the key to a resolution of this problem is already contained in the seminal work of GKP [63] discussed above. They show explicitly (based on earlier work) that the metric on a CY orientifold threaded by 3-form-flux is not of product type but **warped**:

$$ds^2 = \Omega^2(y)\eta_{\mu\nu}dx^\mu dx^\nu + g_{mn}(y)dy^m dy^n\,. \tag{6.54}$$

Here x^μ (with $\mu = 0, \cdots, 3$) and y^m (with $m = 1 \cdots 6$) parameterize the non-compact $\mathbb{R}^4$ and compact X_6 part of our total space respectively. This space is, topologically and as a differentiable manifold, still of product type, $\mathbb{R}^4 \times X_6$. However, the metric manifold built on this basis does not share this product structure. As we can see from the warped metric ansatz in (6.54), this breaking of the product structure is perfectly consistent with 4d Poincare invariance as long as y enters in the prefactor of the non-compact metric but x does *not* enter in the prefactor of the compact part of the metric. One refers to $\Omega(y)$ as the **warp factor**.

Moreover, GKP show that given certain (very common[16]) features of the CY and a particular flux choice, the compact manifold develops a strongly warped region. This region is also known as a **Klebanov-Strassler throat** [90] and is graphically often represented as in Fig. 31. To understand that the compact geometry is strongly deformed at strong warping, one also needs to know that [63, 90]

$$g_{mn}(y) = \Omega^{-2}\tilde{g}_{mn}(y)\,, \tag{6.55}$$

where $\tilde{g}$ is the Calabi-Yau metric. But this is not essential for us. What *is* essential is energetic effect of the warping on the SUSY-breaking $\overline{D3}$ brane placed in the Calabi-Yau orientifold.

To understand this central aspect, recall the Schwarzschild black hole metric

$$ds^2 = -f(r)dt^2 + f^{-1}(r)dr^2 + r^2 d\omega^2\,, \tag{6.56}$$

[16]The feature we need is a so-called conifold singularity. The latter develops when a a certain type of 3-cycle shrinks to zero volume (i.e. $z \to 0$ if the z is the modulus parameterizing the corresponding period). This is in fact a generic type of 3-cycle of a CY, so such a situation arises frequently. Conversely, the conifold singularity can be made smooth ('deformed') by 'blowing up' a 3-cycle. For more details see e.g. [91].

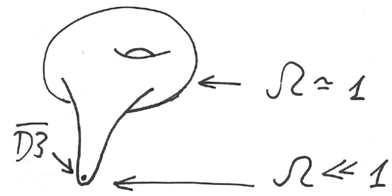

Figure 31: Calabi Yau with warped throat. The Calabi-Yau is basically undeformed in the region of small warping ($\Omega \simeq 1$) and strongly deformed in the $\Omega \ll 1$ domain (where the $\overline{D3}$ brane is localized).

where $d\omega^2$ is the metric on the unit sphere. Clearly, $f(r)$ bears similarity to our $\Omega^2(y)$. As is well known, the vanishing of $f(r)$ as one approaches the horizon is responsible for the redshift effect and the force that pulls any massive object into the black hole. The same happens in our case: The $\overline{D3}$ brane represents a SUSY-breaking local energy density in the warped Calabi-Yau and this brane is pulled towards strong warping (where $\Omega \ll 1$). Once there, its energetic effect as seen from the unwarped 'bulk' of the Calabi-Yau is greatly reduced. In other words, the anti-brane naturally sits at the bottom of the warped throat and uplifts the total potential energy of the compactification only by

$$\Omega^4_{min} \times \mathcal{O}(1) \tag{6.57}$$

in string units. The fourth power of Ω arises since, as known from black hole physics, $f^{1/2}$ is the redshift factor and, in our context, we are redshifting an energy density, i.e. an object of mass dimension four.

As shown in GKP,

$$\Omega_{min} \sim \exp(-2\pi K/3Mg_s)\,, \tag{6.58}$$

where K and M are flux numbers associated with 3-cycles of the KS-throat geometry and g_s is the string coupling constant. The latter is governed by the modulus S stabilized by fluxes. Thus, one apparently has enough freedom to choose fluxes in such a way that Ω_{min} is exponentially small.[17]

Before moving on, it should be mentioned that a debate about the metastability of the anti-brane at the bottom of the throat has been going on for a number of years (see [95–100] and refs. therein). Indeed, as should be clear form the above the $\overline{\text{D3}}$ breaks SUSY (in the absence of any D3 brane) against the fluxes in the throat. It can annihilate against these fluxes only at the price of overcoming an energy barrier, making the uplifted configuration at best metastable [94]. However, the backreaction of fluxes to the presence of the anti-brane is poorly understood and a barrier-free decay or outright instability has been claimed. Yet, in my understanding, metastability as described in [94] has remained plausible [96, 98]. On the other hand, a better, fully-backreacted understanding of the geometry with the anti-brane included would be highly desirable but remains challenging.

Let us now assume that the above $\overline{\text{D3}}$ uplift does indeed provide metastable SUSY breaking and estimate its magnitude. For simplicity, we disregard factors of g_s such that the tension of 3-brane in either the 10d string or Einstein frame is $\mathcal{O}(1) \times l_s^{-4} \sim \mathcal{O}(1)$. Here we also, as before, use

[17]See [92, 93] for recent, in part critical comments related to this point.

conventions in which all dimensionful quantities in 10d are measured in units of the string scale or the inverse string scale. If there were no warping then, compactifying, the $\overline{\text{D3}}$ brane tension (more precisely twice this number - see above) induces a 4d energy density $\sim \mathcal{O}(1)$. Note that we are still using string units and our 4d Planck mass is $M_P^2 \sim \mathcal{V}$ (i.e. we are in a 'Brans-Dicke frame'). Next, we Weyl rescale the 4d metric to go to the 4d Einstein frame. This amounts to using 4d Planck units (i.e. setting the Planck mass to unity) in the 4d effective action. Since, in this process, dimensionless ratios of physical observables do not change, we have

$$\frac{\rho_{\overline{\text{D3}}}^{Eistein}}{(M_P^4)_{Einstein}} \sim \frac{\rho_{\overline{\text{D3}}}^{Brans-Dicke}}{(M_P^4)_{Brans-Dicke}} \sim \frac{1}{\mathcal{V}^2} \qquad \text{or} \qquad \rho_{\overline{\text{D3}}}^{Eistein} \sim \frac{1}{\mathcal{V}^2} \sim \frac{1}{\tau^3}\,. \tag{6.59}$$

Most naively, one would now like to include warping by multiplying with the fourth power of the redshift factor Ω_{min}^4 [86]. This is correct in principle, but at a quantitative level a further fine point has to be taken into account [101]: Indeed, the expression (6.58) is valid in the strongly warped region near the tip of a Klebanov-Strassler throat. It represents correctly the dependence of the warping on the relevant discrete flux choice. Yet, if the the Calabi-Yau volume is taken to infinity, then it is clear that eventually the fluxes become so diluted that their backreaction on the geometry is negligible and $\Omega \sim 1$, even at the lowest point of the throat. This can be quantified [101] and leads to the more precise warping suppression

$$\Omega_{min}^4 \quad \rightarrow \quad \Omega_{min}^4 \tau\,, \tag{6.60}$$

valid only as long as $\Omega_{min}^4 \tau \ll 1$.

Combining everything, one arrives at

$$V_{KKLT} = e^K \left(K^{T\overline{T}} |D_T W|^2 - 3|W|^2 \right) + V_{up}(\tau)\,, \tag{6.61}$$

with

$$K = -3\ln(T + \overline{T})\,, \qquad W = W_0 + Ae^{-aT} \qquad \text{and} \qquad V_{up}(\tau) = c\frac{\Omega_{min}^4}{\tau^2}\,. \tag{6.62}$$

Here A, a and c are numerical $\mathcal{O}(1)$ factors and W_0 and Ω_{min} can be chosen extremely small by an appropriate flux choice. It is easy to convince oneself numerically or analytically that an uplifted situation with a metastable dS or Minkowski vacuum as in Fig. 29 can be achived on the basis of the above potential. The reader is invited to verify this. They key non-trivial point is that the AdS minimum is very steep (based on the exponential behavior of the non-perturbative superpotential $\sim e^{-aT}$) while the uplift has a relatively flat, power-like τ dependence. Hence, the local minimum survives the uplift to a value above zero.[18]

[18]We note in passing that a new round of criticism and defense of this construction has appeared relatively recently, related mainly to the question whether the non-perturbative effect (in this case gaugino condensate) and the subsequent uplift can also be understood directly in 10d [102–107]. At this point it appears that, yet again, the success of the KKLT construction remains plausible [105]. An interesting novel criticism raised in [107] concerns the achievable size of the bulk CY and the question whether this size is large enough for semiclassical control.

6.6 The Large Volume Scenario

A very promising alternative to the KKLT proposal for Kahler moduli stabilization in an AdS vacuum (before uplift) is provided by the Large Volume Scenario or LVS [108]. It has the disadvantage of being slightly more involved than KKLT but the advantage that the stabilized value of the volume $\mathcal{V}_{LVS}$ may be *exponentially* large – a feature not available in KKLT due to the parametric behavior $\mathcal{V}_{KKLT} \sim \ln(1/|W_0|)$.

In the simplest realization, two Kahler moduli T_b and T_s (with the indices standing for 'big' and 'small') are required. The volume is assumed to take the form

$$\mathcal{V}(\tau_b, \tau_s) \;\sim\; \tau_b^{3/2} - c\tau_s^{3/2}\,, \tag{6.63}$$

with $2\tau_i = T_i + \overline{T}_i$ and $K_K = -2\ln\mathcal{V}$, as usual.

...to be extended...

6.7 Vacuum Statistics and the realizability fine tunings

Let us now assume that one of the moduli stabilization and uplifting procedures discussed in the literature (the two main examples being KKLT and LVS) or some variant thereof works. This implies the existence of a landscape of 4d EFTs with a certain random distribution of operator coefficients, including in particular the cosmological constant λ and the Higgs mass parameter m_H^2. Two non-trivial questions can then be asked. First, is it clear that the landscape contains a vacuum with the apparently highly fine-tuned values of λ and m_H^2 we observe? Second, can we understand why we find ourselves in an world described by such a very special vacuum?

In this section, we want to discuss, at least briefly, the first (and simpler) of these two questions. We focus on λ and on KKLT. In this case, a partial answer can be given using a fundamental technical result of [87] (see also [45, 109, 110]). In this paper the focus is entirely on the flux stablization of complex structure moduli (and the axio-dilaton), i.e., Kahler moduli are ignored. The setting is (for our purposes) that of type IIB Calabi-Yau orientifolds with O3/O7-planes.[19] In this setting, the tadpole constraint on the flux vector (f, h) can be calculated (for details see below) such that one knows precisely in which subset of the space of integer vectors this object takes its values. Each such value corresponds to a point in complex-structure moduli space at which the geometry (the variables z_i and S) is then stabilized. If the dimension of the moduli space and hence of the vector (f, h) is large, solving for the z_i on the basis of a given flux value is practically impossible. But, assuming that the set of relevant flux choices is large, it is possible to talk about the resulting (approximately) statistical distribution of vacua in moduli space. In fact, in the strict mathematical limit of a large tadpole (taking the restriction on the length of (f, h) to infinity), this becomes a precise mathematical question.

They key answer given in [87] concerns the distribution of a particular quantity, $e^{K/2}W$. It was shown that, under mild assumptions, the distribution of this number in the complex plane is flat near zero, cf. Fig. 32. We now want to include Kahler moduli (for simplicity a single Kahler

[19] In fact, the proper framework is the non-perturbative generalization of this setting, known as F-theory. See below.

modulus) assuming that, for a large subset of these vacua, instanton or gaugino condensate effects are present. This leads to

$$W_0 \quad \to \quad W_0 + Ae^{-aT}\,, \tag{6.64}$$

eventually giving rise to full moduli stabilization in AdS with

$$\lambda_{AdS} \sim -e^K |W_0|^2\,. \tag{6.65}$$

The flat distribution of the complex number $e^{K/2}W_0$ now implies a flat distribution of λ_{AdS}, reaching up to zero from below (cf. the l.h. side of Fig. 33). After an uplift of the type described in Sect. 6.4 or 6.5, a dense distribution of λ values including the zero-point is obtained (cf. the r.h. side of Fig. 33).

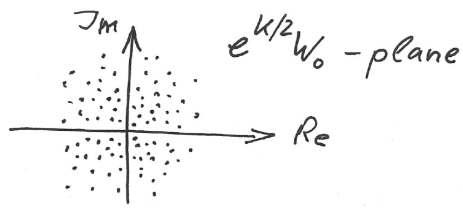

Figure 32: The distribution of $e^{K/2}W_0$ in the complex plane has no special feature near the origin.

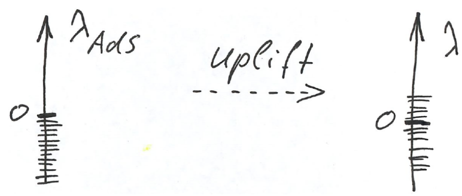

Figure 33: Distribution of the cosmological constant before and after uplift.

It is crucial in this logic that both the value of W_0 and the uplift energy can be extremely small. In the first case, the reason the tuning in the flux discretuum, as described above. In the second case it is the exponential warping suppression. Thus, a value of λ very close to zero can arise after a shallow AdS vacuum is uplifted by a small amount. The restriction to shallow AdS vacua and small uplifts is crucial for calculational control purposes: Specifically, small W_0 implies a relatively large volume and hence a suppression of various higher-order (α' and string loop) corrections.

Of course, it is important to quantify how dense the discretuum is and hence how finely spaced a districution of λ values in Fig. 33 one can hope for. For this, we need to discuss tadpole cancellation for the C_4 potential. By this we mean that the coefficient of the action term linear in C_4 (the 'tadpole') should be zero. The intuition behind this is best explained by an analogy to electrodynamics:

Imagine our space were not $\mathbb{R}^3$ but compact, say S^3. Then by Gauss' law a static solution of the Maxwell equations

$$d * F_2 = d * dA_1 = j_3 \tag{6.66}$$

clearly requires that the total number of sources add up to zero,

$$\int_{S^3} j_3 = 0\,. \tag{6.67}$$

Even more intuitively, the number of electrons and positrons must be the same since there can not be more 'beginnings' than 'ends' of electric field lines on a compact manifold.

In our case, the Chern-Simons lagrangian

$$\int C_4 \wedge F_3 \wedge H_3 \equiv \int C_4 \wedge j_6^{flux} \tag{6.68}$$

implies that part of the sources for C_4 are provided by the 3-form flux. Moreover, the type IIB equations of motion also imply that G_3 is imaginary self-dual,

$$*G_3 = iG_3\,, \tag{6.69}$$

which in turn implies that $\int_{CY} j_6^{flux}$ can not get different-sign contributions from different regions of the CY. The contribution of the fluxes to the C_4 source or the so-called 'D3 tadpole' can be written as

$$\int_{CY} j_6^{flux} = \int F_3 \wedge H_3 \sim (h, f)^2\,, \tag{6.70}$$

with an appropriately defined (symplectic) product on the space of flux vectors (h, f).

This flux vector contribution to the D3 tadpole has to be cancelled by other charged objects. We are discussing this before the uplift, so $\overline{\text{D3}}$ branes are not at our disposal. D3 branes contribute with the same sign as (supersymmetric) 3-form fluxes. The available options are then only O3-planes or O7-planes. The first contribute in an obvious way since they are charged oppositely w.r.t. D3 branes. By contrast, the O7-brane contribtuion is indirect and involves an integral over the curvature of the O7-plane, which can be viewed as a curved, co-dimension-2 object in the 6d CY. We are not going to spell this out explicitly, but only report the results of a more general analysis:

Type IIB compactifications in the perturbative regime find their non-perturbative completion in so-called F-theory models [111] (for reviews see [1, 112, 113]). Their detailed discussion goes beyond the scope of this subsection. Suffice it to say that these are based on the geometry of an elliptically-fibred (roughly torus-fibred) CY 4-fold. The fibre torus encodes the information that corresponds, in type IIB language, to the variation of the axio-dilaton S. In fact, S is identified with the complex-structure parameter of the fibre torus.

In this more general F-theory setting, the tadpole contribution of the O7-planes above is encoded in the 4-fold geometry, more precisely in the Euler-number χ_4 of the 4-fold. But this is much more general - contributions arise also from 7-branes other than the standard D7-branes of perturbative type IIB string theory. In our context, the crucial constraint then becomes

$$N^T \Sigma N \leq L \equiv \frac{\chi_4}{24} \qquad \text{with} \qquad N = \begin{pmatrix} h \\ f \end{pmatrix} \qquad \text{and} \qquad \Sigma \equiv \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{6.71}$$

The key geometric input is the availability of 4-folds with Euler characteristics up to $\chi_4 \sim 10^6$ (see e.g. [114]), leading to $L \sim 10^5$. The number of vacua can then be estimated as [87]

$$\mathcal{N}_{vac} \sim \frac{L^K}{K!}\,, \tag{6.72}$$

where K is the number of 3-cycles of the CY. This number is crucial in the present context since it determines the dimension of the flux vector N to be $d = 2K$. Thus, the estimate of $\mathcal{N}_{vac}$ above can roughly be understood as the volume of a $2K$-dimensional ball of radius $\sqrt{L}$. This is a natural expectation since we are dealing with a lattice with unit spacing on which the flux vector can end. Of this lattice only a certain subset, specified by the inequality in (6.71), is available. The details are slightly more complicated since the metric Σ is not positive definite, such that the 'radius' $\sqrt{L}$ does not specify a ball but the interior of a hyperboloid (the non-compact directions of which are, however, cut off by physical arguments and do not lead to a divergence of $\mathcal{N}_{vac}$).

In the end, using the (far from maximal) numbers $L = 10^4$ for the 4-fold Euler number and $K = 300$ for the number of 3-cycles of the corresponding Calabi-Yau orientifold, one arrives at

$$\mathcal{N}_{vac} \sim (eL/K)^K \sim 100^{300} = 10^{600}\,. \tag{6.73}$$

Even after appropriate reductions for the geometric constraints implied by the gaugino condensate / instanton effect and by the warped throat required for the uplift, this is still more than sufficient to realize the desired fine tuning for the cosmological constant $\sim 10^{-120}$ (in Planck units). In fact, most naively (ignoring the reduction by geometric constraints) one expects that of the 10^{600} vacua about 10^{480} have a cosmological constant of the order of 10^{-120}.

At this point, a comment concerning a more recent development in the context of vacuum counting has to be made. It concerns the number of several hundred 3-cycles which we used and which is typical for a CY 3-fold. Clearly, an O7-orientifold of a CY 3-fold has more moduli due to the freedom of deforming the 4 D7 branes that originally lie on top of each O7 plane. Even more generally, similar situations can be analysed in the F-theory context, where more types of co-dimension-2 objects than just O7-branes and D7-planes are available. In this context, the 3-fold complex structure and D7-brane deformation moduli are unified as complex-structure moduli of the elliptically fibred CY 4-fold. In this F-theory setting, 'the geometry with most flux vacua' has recently been identified [115]. The tadpole is similar to what was discussed above, but the relevant number of 4-fold complex structure moduli is $h^{3,1} = 303,148$. This leads to $\mathcal{O}(10^{272,000})$ flux vacua – far beyond what can be expected on the basis of just 3-fold complex structure moduli.

Assuming that Higgs-like scalars with a flat (flux-related) mass-squared distribution going through zero are available, one can discuss the fine tuning of the Higgs mass parameter analogously to the tuning of the cosmological constant above. Up to model building constraints, the reduction in the number of vacua by this requirement is by a factor of $(100\,\mathrm{GeV})^2/(10^{18}\,GeV)^2 \sim 10^{-32}$. Thus, we would apparently still be left with $\sim 10^{450}$ vacua with accidentally small cosmological constant *and* a light Higgs.

Needless to say, the problem of tuning the Higgs mass small is alleviated if also have low-scale (or at least relatively low-scale) SUSY. Such models with low-scale SUSY are also available in the landscape, in the simplest case by tuning W_0, which determines $m_{3/2}$, to be sufficiently small. An interesting question is now whether we are more likely to find ourselves in a world with purly fine-tuned light Higgs or with a light Higgs mostly due to SUSY (possibly with some extra tuning in addition). One part of the answer can be given by asking how many of the respective vacua are available. In other words, is it 'cheaper' to directly tune for a small Higgs mass or to tune for a low SUSY breaking scale. The second option looks advantageous since there are also model-building possibilities to lower the SUSY-breaking scale. Yet another option would be to look for models with technicolor-like structure, lowering the Higgs mass in a non-SUSY-related dynamical way.

Yet, vacuum counting alone is not sufficient to settle the interesting questions above. Indeed, it is possible that many more vacua with low-scale SUSY rather than with purely fine-tuned non-SUSY light Higgs are available. But this would become irrelevant if cosmological dynamics prefers inflation to always end in vacua with high-scale SUSY-breaking. Thus, we need to turn to the dynamics which might be responsible for populating the landscape.

7 Eternal Inflation and the Measure Problem

7.1 From slow-roll inflation to the eternal regime

The present course does, of course, assume General Relativity as a prerequisite. Since most relativity courses include some cosmology, it appears logical to assume that the reader will also be familiar with the most basic cosmology-related formulae. We only summarize the results to set our notation:

The cosmological principle, with excellent support from data, postulates that space is homogeneous and isotropic on large scales. Together, these two features imply spacetime can be represented as a 1-parameter family of spatial homogeneous hypersurfaces H_t (with $t \in \mathbb{R}$) which are threaded orthogonally by 'observer curves'. Each of those is parameterized by the oberserver eigentime t. In terms of the metric, this means

$$ds^2 = -dt^2 + a^2(t)\, g_{ij}\, dx^i dx^j\,, \tag{7.1}$$

where a is the scale factor and g_{ij} is metric on a maximally symmetric 3d space, i.e. the a sphere, a plane or 3d hyperboloid.

If matter comes in the form of a perfect fluid with

$$T_{\mu\nu} = \rho\, u_\mu u_\nu + p\,(g_{\mu\nu} + u_\mu u_\nu)\,, \tag{7.2}$$

with density ρ and pressure p, then the Einstein equations and the coninuity equation reduce to

$$3M_P^2(H^2 + 3k^2/a^2) \;=\; \rho \tag{7.3}$$
$$\dot\rho + 3H\,(\rho + p) \;=\; 0\,, \qquad \text{with the Hubble parameter} \qquad H = \dot a/a\,. \tag{7.4}$$

Here $k = +1, 0, -1$ distinguishes the three cases of positive, zero and negative spatial curvature.

A case of particular interest is that of a scalar ϕ with potential $V(\phi)$. Using the standard result $\rho = T + V$ and $p = T - V$ (with $T = \dot\phi^2/2$), one then immediately sees that (7.4) takes the form

$$\ddot\phi + 3H\dot\phi + V' = 0\,. \tag{7.5}$$

Standard slow-roll inflation arises in the regime where the potential V is sufficiently flat. This is conventionally quantified by requiring smallness of the two slow-roll parameters ($M_P = 1$ here and below):

$$\epsilon \equiv \frac{1}{2}\left(\frac{V'}{V}\right)^2 \ll 1 \qquad \text{and} \qquad \eta \equiv \frac{V''}{V} \ll 1\,. \tag{7.6}$$

Indeed, in this regime $\ddot{\phi}$ can be neglected in the equation of motion for ϕ and ρ is dominated by the potential energy. Thus, cosmology is described by

$$3H\dot{\phi} = -V' \qquad \text{with} \qquad H^2 = V/3 \qquad \text{and} \qquad a = \exp(Ht)\,. \tag{7.7}$$

This represents a so-called quasi-de-Sitter situation, exact de-Sitter expansion corresponding to an exactly constant (rather than slowly changing) H in the last equation of (7.7).

In standard cosmology one assumes that this situation lasts long enough to explain the flatness and homogeneity of our present-day universe. But eventually it ends since ϕ rolls into a region where the slow-roll conditions cease to hold, ϕ oscillates about its minimum and eventually decays to Standard Model particles, reheating the universe (cf. Fig. 34).

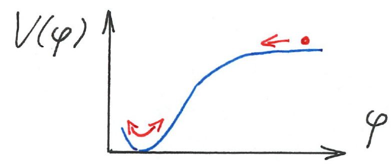

Figure 34: Slow-roll inflation ending in field oscillations and reheating.

Crucially, while in the slow-roll regime, ϕ does not only roll classically but is, at the same time, subject to quantum fluctuations. To understand this qualitatively, it is useful to consider the simplified case of pure de Sitter with a an exactly massless scalar ((V = const. and hence H = const.). It is then easy to determine the inward-going geodesics in the relevant metric ($k = 0$ for simplicity)

$$ds^2 = -dt^2 + e^{2Ht}d\overline{x}^2\,. \tag{7.8}$$

One finds that, above some maximal radius r_0 (with $r^2 \equiv d\overline{x}^2$), they never reach the origin. In other words, there exists a cosmological horizon, of the order of the so-called de sitter radius $1/H$. Each spatial slice falls into many s-called de Sitter patches, which are causally disconnected. As the universe evolves, the exponential expansion roughly doubles the number of those once in a Hubble time (which is also $1/H$).

7.2 Slow-roll inflation in the landscape

A very nice and pedagogical review is [116]. More references will be given below.

Consider a scalar field theory with a potential V and several local minima (vacua). Each of them has a different cosmological constants $\lambda = V(\phi_{min})$. If at least one of those minima has $\lambda > 0$ and if the probability T for tunneling out of this minimum (per volume and time) is smaller than the fourth power of its expansion rate,

$$T \lesssim H^4(\lambda)\,, \qquad \text{where} \qquad 3M_P^2 H^2(\lambda) = \lambda\,, \tag{7.9}$$

this already gives rise to eternal inflation, cf. the schematic (Penrose-type) diagram in Fig. 35. Condition (7.9) may roughly be understood as follows: The density (in 4-volume) of nucleation points of bubbles of other vacua is small on the typical scale H of the underlying dS space. There

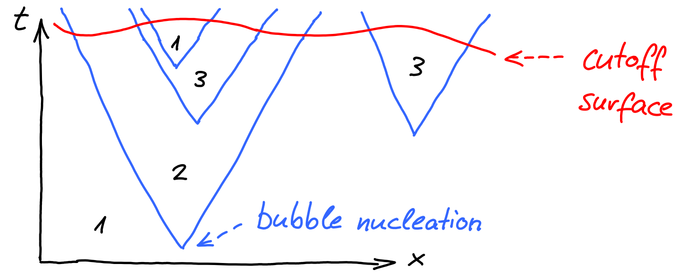

Figure 35: Nucleation and speed-of-light expansion of bubbles in a 'background' dS vacuum. The 'cutoff surface' will be discussed later.

is then no danger that the loss of volume to other vacua wins over the volume growth due to de Sitter expansion.

In the generic case there are of course more than one dS vacuum. There is then both down and up-tunneling (even though the latter is strongly suppressed), cf. Fig 36. Thus, the whole landscape gets populated once eternal inflation is running. The non-dS vacua are called 'terminal' since there is no way back from them to the dS part of the landscape. The only possible future of an observer in one of them is a big crunch or an eternal, stable Minkowski space.

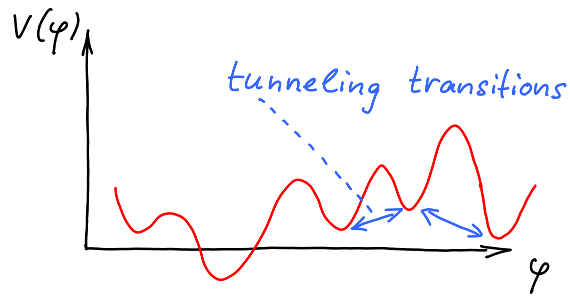

Figure 36: Down and up-tunneling in the landscape.

In string theory, the scalar manifold on which our potential with the various minima lives is, in the simplest case, the moduli space of a certain Calabi-Yau (or Calabi-Yau orientifold). To be precise, one has to allow for gluing the moduli spaces of many different compact geometries, with the geometric transitions between them (in many cases well understood, in other cases conjectural) also being realised by tunneling.

As an important side-remark, let us note that the picture of a smooth potential over the moduli space is an oversimplification. The potential comes from fluxes (discrete expectation values of p-form field strengths) and does indeed have smooth minima. But the transition between two different minima usually involves passing a domain wall (which microscopically corresponds to a D-brane). In other words, the 'saddles' connecting different minima in the 'potential landscape picture' we started with are far from being smooth. But this is not essential for what follows.

7.3 The measure problem

Accepting the above landscape picture and eternal inflation as the process populating it, the measure problem is easy to state, at least at an intuitive level: We live in one of the vacua, but we do not know in which one. We would like to make a statistical prediction (given that we know certain features of our vacuum, but not all). Let us say the new observable which we are going to measure tomorrow can take the values A or B. The most naive way to make a statistical prediction would be to say that the ratio of probabilities is

$$p_A/p_B = N_A/N_B\,. \tag{7.10}$$

Here $N_{A/B}$ are the numbers of observers in the multivers who have measured all that we have measured so far and who will, in the next measurement, find A or B respectively. But in eternal inflation, by definition, both numbers are infinite and their ratio is not well-defined. What is worse, if one cuts off the infinity in the future (e.g. by restricting attention to measurements before some maximal time t_{max}, with a limiting procedure $t_{max} \to \infty$ in the end), the prediction becomes dependent on the precise type of cutoff. In the case of a maximal time, this is due to the absence of a unique global time variable in the geometry of Fig. 35.

A word concerning our place in the above eternal inflation picture might be useful. Naively one might think that we live in one of the many bubbles, and ours just happens to have very small λ. This is roughly true, but important details are missing. First, given how small our λ is, we naturally expect the previous vacuum's λ to be much larger. But a corresponding tunneling event would have endowed our vacuum with a large and negative spatial curvature. Our cosmological evolution would have been governed by the FRW equation

$$3M_P^2H^2 = \rho - 3M_P^2k/a^2 \qquad \text{with} \qquad k=-1 \tag{7.11}$$

and with initital conditions where the curvature term (the 2nd term on the r.h. side) would be at least comparable to the matter term (ρ, which includes matter, radiation and λ) from the start. In such a situation, there can be either λ domination or curvature domination succeeded by λ domination, but no extended radiation or matter dominated epoch, as in our world. The reason is simply that, with expansion, matter and radiation densities decay faster than curvature.

The way out is to postulate that our local minimum has the peculiar feature of an inflationary plateau (it does not actually have to be a plateau - any sufficiently flat potential region would do) where cosmological inflation took place, diluting the curvature contribution (see Fig. 37). Thus, observers of 'our kind' actually always appear shortly after the tunneling transition. Nevertheless, in a given bubble of 'our vacuum' their number is infinite, as are our reheating and structure formation surface (see Fig. 38). These surfaces (including presumably the surface of death of all stars and hence of all civilzations) follow the straight bubble wall surface all the way up to infinity (both left and right, assuming there is no bubble collision). One may say that the interior of any bubble is an open (infinite) FRW universe.

References

[1] F. Denef, "Les Houches Lectures on Constructing String Vacua," Les Houches **87** (2008) 483 [arXiv:0803.1194].

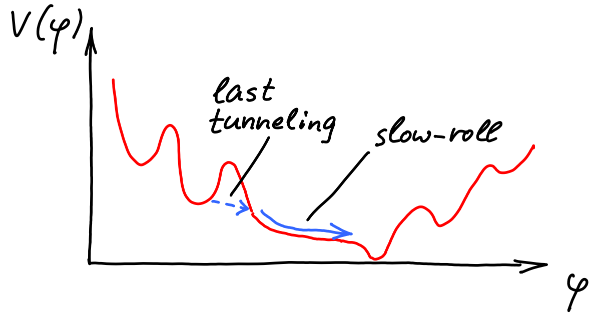

Figure 37: Tunneling to our vacuum, where a period of slow-roll inflation, reheating and structure formation preceed the dS phase.

[2] Ibanez/Uranga: String Theory and Particle Physics.

[3] Peski/Schroeder: An Introduction to Quantum Field Theory.

[4] Weinberg: The Quantum Theory of Fields, Vol. I and II.

[5] Itzykson/Zuber: Quantum Field Theory.

[6] Srednicki: Quantum Field Theory.

[7] M.D. Schwartz: Quantum Field Theory and the Standard Model.

[8] Cheng/Li: Gauge Theory of Elementary Particle Physics.

[9] Cheng/Li: Gauge Theory of Elementary Particle Physics – Problems and Solutions.

[10] Donoghue/Golowich/Holstein: Dynamics of the Standard Model.

[11] Nachtmann: Elementary Particle Physics.

[12] M. Blanke, "Introduction to Flavour Physics and CP Violation," CERN Yellow Rep. School Proc. **1705** (2017) 71 [arXiv:1704.03753 [hep-ph]].

[13] Kooijman/Tuning: Lectures on CP violation, https://www.nikhef.nl/ h71/talks/talks.html

[14] A. A. Anselm and A. A. Johansen, "Can electroweak theta term be observable?," Nucl. Phys. B **412** (1994) 553 [hep-ph/9305271].

[15] C. Cao and A. Zhitnitsky, "Axion detection via Topological Casimir Effect," Phys. Rev. D **96** (2017) no.1, 015013 [arXiv:1702.00012 [hep-ph]].

[16] S. D. H. Hsu, "Theta terms and asymptotic behavior of gauge potentials in (3+1) dimensions," arXiv:1107.0756 [hep-th].

[17] M. Luty, "2004 TASI lectures on supersymmetry breaking," hep-th/0509029.

[18] D. B. Kaplan, "Five lectures on effective field theory," nucl-th/0510023.

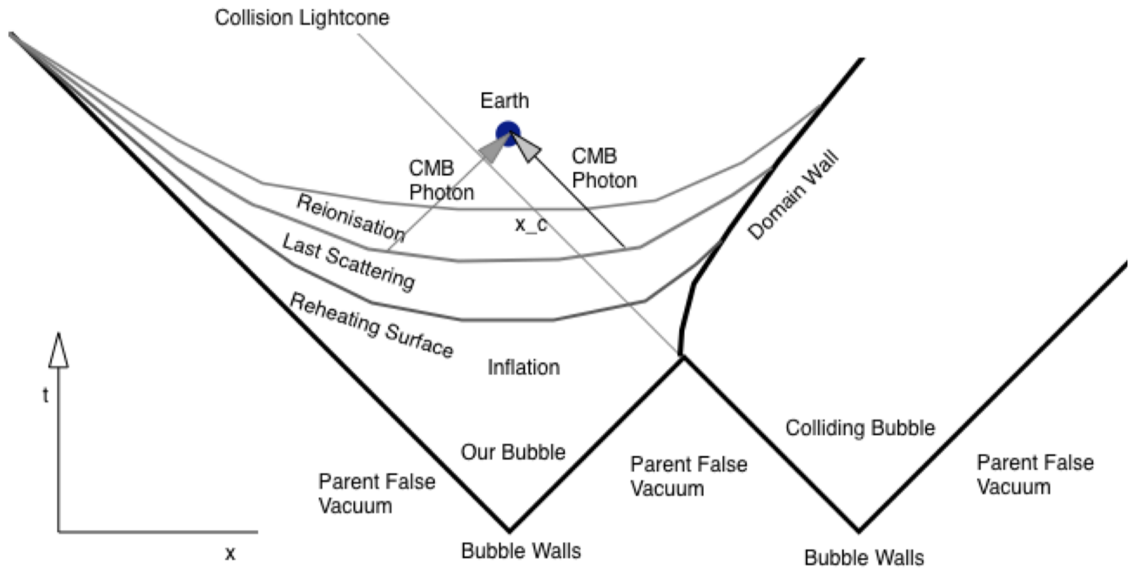

Figure 38: Various 'surfaces of constant energy density' following the initial tunneling transition to our bubble. The picture is taken from [120], which deals with possible bubble collisions. That aspect of the figure is not important for us right now.

[19] Minkowski '77; Yanagida, Gell-Mann/Ramond/Slansky '79.

[20] Schellekens: Beyond the Standard Model, `https://www.nikhef.nl/ t58/lectures.html`

[21] A. N. Schellekens, "The String Theory Landscape," Adv. Ser. Direct. High Energy Phys. **22** (2015) 155.

[22] G. 't Hooft, C. Itzykson, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer and R. Stora, "Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979," NATO Sci. Ser. B **59** (1980) pp.1.

[23] R. Barbieri and G. F. Giudice, "Upper Bounds on Supersymmetric Particle Masses," Nucl. Phys. B **306** (1988) 63.

[24] J. R. Ellis, K. Enqvist, D. V. Nanopoulos and F. Zwirner, "Observables in Low-Energy Superstring Models," Mod. Phys. Lett. A **1** (1986) 57.

[25] J. D. Wells, "Naturalness, Extra-Empirical Theory Assessments, and the Implications of Skepticism," arXiv:1806.07289 [physics.hist-ph] and "Finetuned Cancellations and Improbable Theories," arXiv:1809.03374 [physics.hist-ph].

[26] M. Tanabashi *et al.* [Particle Data Group], "Review of Particle Physics," Phys. Rev. D **98** (2018) no.3, 030001.

[27] Wald: General Relativity.

[28] Weinberg: Gravitation and Cosmology.

[29] Misner/Thorne/Wheeler: Gravitation.

[30] Carroll: Spacetime and Geometry.

[31] Straumann: General Relativity.

[32] M. Veltman, "Quantum theory of gravitation," in: *Methods in Field Theory, Les Houches 1975*, edited by R. Balian and J. Zinn-Justin (North-Holland Publ., Amsterdam, the Netherlands, 1976), Course 5.

[33] S. R. Coleman and J. Mandula, "All Possible Symmetries of the S Matrix," Phys. Rev. **159** (1967) 1251.

[34] R. Haag, J. T. Lopuszanski and M. Sohnius, "All Possible Generators of Supersymmetries of the s Matrix," Nucl. Phys. B **88** (1975) 257. doi:10.1016/0550-3213(75)90279-5

[35] Wess/Bagger: Supersymmetry and Supergravity.

[36] West: Introduction to Supersymmetry and Supergravity.

[37] Weinberg: Quantum Field Theory, Vol. 3.

[38] Turning: Modern Supersymmetry.

[39] Shifman: Advanced Topics in Quantum Field Theory.

[40] N. Seiberg, "Naturalness versus supersymmetric nonrenormalization theorems," Phys. Lett. B **318** (1993) 469 [hep-ph/9309335].

[41] S. P. Martin, "A Supersymmetry primer," Adv. Ser. Direct. High Energy Phys. **21** (2010) 1 [hep-ph/9709356].

[42] G. F. Giudice and R. Rattazzi, "Theories with gauge mediated supersymmetry breaking," Phys. Rept. **322** (1999) 419 [hep-ph/9801271].

[43] N. Arkani-Hamed and S. Dimopoulos, "Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC," JHEP **0506** (2005) 073 [hep-th/0405159].

[44] G. F. Giudice and A. Romanino, "Split supersymmetry," Nucl. Phys. B **699** (2004) 65 Erratum: [Nucl. Phys. B **706** (2005) 487] [hep-ph/0406088].

[45] F. Denef and M. R. Douglas, "Distributions of nonsupersymmetric flux vacua," JHEP **0503** (2005) 061 doi:10.1088/1126-6708/2005/03/061 [hep-th/0411183].

[46] A. Hebecker, A. K. Knochel and T. Weigand, "A Shift Symmetry in the Higgs Sector: Experimental Hints and Stringy Realizations," JHEP **1206** (2012) 093 [arXiv:1204.2551 [hep-th]].

[47] Buchbinder/Kuzenko: Ideas and Methods of Supersymmetry and Supergravity.

[48] F. Quevedo, S. Krippendorf and O. Schlotterer, "Cambridge Lectures on Supersymmetry and Extra Dimensions," arXiv:1011.1491 [hep-th].

[49] Freedman/Van Proeyen: Supergravity.

[50] G. Villadoro and F. Zwirner, "De-Sitter vacua via consistent D-terms," Phys. Rev. Lett. **95** (2005) 231602 [hep-th/0508167].

[51] Green/Schwarz/Witten: Superstring theory, vol. I and II.

[52] Polchinski: String theory, vol. I and II.

[53] Blumenhagen/Lüst/Theisen: Basic concepts of string theory.

[54] Becker/Becker/Schwarz: String theory and M-theory.

[55] Kiritsis: String theory in a nutshell.

[56] Zwiebach: A first course in string theory.

[57] A. A. Belavin, A. M. Polyakov and A. B. Zamolodchikov, "Infinite Conformal Symmetry in Two-Dimensional Quantum Field Theory," Nucl. Phys. B **241** (1984) 333.

[58] Francesco/Mathieu/Senechal: Conformal Field Theory.

[59] Schottenloher: A Mathematical Introduction to Conformal Field Theory.

[60] Blumenhagen/Plauschinn: Introduction to Conformal Field Theory.

[61] P. H. Ginsparg, "Applied Conformal Field Theory," Lectures at Les Houches Summer School 1988, hep-th/9108028.

[62] A.N. Schellekens, Conformal Field Theory, lecture notes, `https://www.nikhef.nl/ t58/Site/Lectures.html`

[63] S. B. Giddings, S. Kachru and J. Polchinski, "Hierarchies from fluxes in string compactifications," Phys. Rev. D **66** (2002) 106006 [hep-th/0105097].

[64] G. Nordstrom, "On the possibility of unifying the electromagnetic and the gravitational fields," Phys. Z. **15** (1914) 504 [physics/0702221 [physics.gen-ph]].

[65] O. Klein, "Quantum Theory and Five-Dimensional Theory of Relativity." Z. Phys. **37** (1926) 895 [Surveys High Energ. Phys. **5** (1986) 241].

[66] O. Klein, "The Atomicity of Electricity as a Quantum Theory Law," Nature **118** (1926) 516.

[67] T. Appelquist, A. Chodos and P. G. O. Freund, "Modern Kaluza-klein Theories,", FRONTIERS IN PHYSICS, 65 (1987).

[68] M. J. Duff, "Kaluza-Klein theory in perspective," talk at The Oskar Klein Centenary, Stockholm, 1994, hep-th/9410046.

[69] J. M. Overduin and P. S. Wesson, "Kaluza-Klein gravity," Phys. Rept. **283** (1997) 303 [gr-qc/9805018].

[70] B. R. Greene, "String theory on Calabi-Yau manifolds," lectures at TASI 1996, hep-th/9702155.

[71] P. Candelas, "Lectures On Complex Manifolds," in "Trieste 1987, Proceedings, Superstrings 1987", pp. 1-88.

[72] T. Hübsch: Calabi-Yau manifolds: A Bestiary for physicists, World Scientific, 1991.

[73] P. Candelas and X. de la Ossa, "Moduli Space of Calabi-Yau Manifolds," Nucl. Phys. B **355** (1991) 455;
P. Candelas, "Yukawa Couplings Between (2,1) Forms," Nucl. Phys. B **298** (1988) 458.

[74] A. Font and S. Theisen, "Introduction to string compactification," Lect. Notes Phys. **668** (2005) 101.

[75] T. W. Grimm and J. Louis, "The Effective action of N = 1 Calabi-Yau orientifolds," Nucl. Phys. B **699** (2004) 387 [hep-th/0403067].

[76] H. Jockers and J. Louis, "The Effective action of D7-branes in N = 1 Calabi-Yau orientifolds," Nucl. Phys. B **705** (2005) 167 [hep-th/0409098].

[77] T. W. Grimm and J. Louis, "The Effective action of type IIA Calabi-Yau orientifolds," Nucl. Phys. B **718** (2005) 153 [hep-th/0412277].

[78] M. Kerstan and T. Weigand, "The Effective action of D6-branes in N=1 type IIA orientifolds," JHEP **1106** (2011) 105 [arXiv:1104.2329 [hep-th]].

[79] A. Giryavets, "New attractors and area codes," JHEP **0603** (2006) 020 [hep-th/0511215].

[80] A. Hebecker, T. Mikhail and P. Soler, "Euclidean wormholes, baby universes, and their impact on particle physics and cosmology," arXiv:1807.00824 [hep-th].

[81] Nakahara: Geometry, Topology and Physics.

[82] Goeckeler/Schuecker: Differential Geometry, Gauge Theories, and Gravity.

[83] R. Bousso and J. Polchinski, "Quantization of four form fluxes and dynamical neutralization of the cosmological constant," JHEP **0006** (2000) 006 [hep-th/0004134].

[84] S. Gukov, C. Vafa and E. Witten, "CFT's from Calabi-Yau four folds," Nucl. Phys. B **584** (2000) 69 Erratum: [Nucl. Phys. B **608** (2001) 477] [hep-th/9906070].

[85] K. Dasgupta, G. Rajesh and S. Sethi, "M theory, orientifolds and G - flux," JHEP **9908** (1999) 023 [hep-th/9908088].

[86] S. Kachru, R. Kallosh, A. D. Linde and S. P. Trivedi, "De Sitter vacua in string theory," Phys. Rev. D **68** (2003) 046005 doi:10.1103/PhysRevD.68.046005 [hep-th/0301240].

[87] F. Denef and M. R. Douglas, "Distributions of flux vacua," JHEP **0405** (2004) 072 [hep-th/0404116].

[88] U. H. Danielsson and T. Van Riet, "What if string theory has no de Sitter vacua?," Int. J. Mod. Phys. D **27** (2018) no.12, 1830007 [arXiv:1804.01120 [hep-th]].

[89] H. Ooguri, E. Palti, G. Shiu and C. Vafa, "Distance and de Sitter Conjectures on the Swampland," Phys. Lett. B **788** (2019) 180 [arXiv:1810.05506 [hep-th]].

[90] I. R. Klebanov and M. J. Strassler, "Supergravity and a confining gauge theory: Duality cascades and chi SB resolution of naked singularities," JHEP **0008** (2000) 052 [hep-th/0007191].

[91] P. Candelas and X. C. de la Ossa, "Comments on Conifolds," Nucl. Phys. B **342** (1990) 246. doi:10.1016/0550-3213(90)90577-Z

[92] I. Bena, E. Dudas, M. Graña and S. Lüst, "Uplifting Runaways," Fortsch. Phys. **67** (2019) no.1-2, 1800100 [arXiv:1809.06861 [hep-th]].

[93] R. Blumenhagen, D. Kläwer and L. Schlechter, "Swampland Variations on a Theme by KKLT," arXiv:1902.07724 [hep-th].

[94] S. Kachru, J. Pearson and H. L. Verlinde, "Brane / flux annihilation and the string dual of a nonsupersymmetric field theory," JHEP **0206** (2002) 021 [hep-th/0112197].

[95] I. Bena, M. Graña, S. Kuperstein and S. Massai, "Giant Tachyons in the Landscape," JHEP **1502** (2015) 146 [arXiv:1410.7776 [hep-th]].

[96] B. Michel, E. Mintun, J. Polchinski, A. Puhm and P. Saad, "Remarks on brane and antibrane dynamics," JHEP **1509** (2015) 021 [arXiv:1412.5702 [hep-th]].

[97] D. Cohen-Maldonado, J. Diaz, T. van Riet and B. Vercnocke, "Observations on fluxes near anti-branes," JHEP **1601** (2016) 126 [arXiv:1507.01022 [hep-th]].

[98] J. Polchinski, "Brane/antibrane dynamics and KKLT stability," arXiv:1509.05710 [hep-th].

[99] I. Bena, J. Blåbäck and D. Turton, "Loop corrections to the antibrane potential," JHEP **1607** (2016) 132 [arXiv:1602.05959 [hep-th]].

[100] U. H. Danielsson, F. F. Gautason and T. Van Riet, "Unstoppable brane-flux decay of $\overline{\mathrm{D6}}$ branes," JHEP **1703** (2017) 141 [arXiv:1609.06529 [hep-th]].

[101] S. Kachru, R. Kallosh, A. D. Linde, J. M. Maldacena, L. P. McAllister and S. P. Trivedi, "Towards inflation in string theory," JCAP **0310** (2003) 013 [hep-th/0308055].

[102] J. Moritz, A. Retolaza and A. Westphal, "Toward de Sitter space from ten dimensions," Phys. Rev. D **97** (2018) no.4, 046010 [arXiv:1707.08678 [hep-th]].

[103] Y. Hamada, A. Hebecker, G. Shiu and P. Soler, "On brane gaugino condensates in 10d," JHEP **1904** (2019) 008 [arXiv:1812.06097 [hep-th]].

[104] R. Kallosh, "Gaugino Condensation and Geometry of the Perfect Square," Phys. Rev. D **99** (2019) no.6, 066003 [arXiv:1901.02023 [hep-th]].

[105] Y. Hamada, A. Hebecker, G. Shiu and P. Soler, "Understanding KKLT from a 10d perspective," arXiv:1902.01410 [hep-th].

[106] F. F. Gautason, V. Van Hemelryck, T. Van Riet and G. Venken, "A 10d view on the KKLT AdS vacuum and uplifting," arXiv:1902.01415 [hep-th].

[107] F. Carta, J. Moritz and A. Westphal, "Gaugino condensation and small uplifts in KKLT," arXiv:1902.01412 [hep-th].

[108] V. Balasubramanian, P. Berglund, J. P. Conlon and F. Quevedo, "Systematics of moduli stabilisation in Calabi-Yau flux compactifications," JHEP **0503** (2005) 007 [hep-th/0502058].

[109] M. R. Douglas, "The Statistics of string / M theory vacua," JHEP **0305** (2003) 046 [hep-th/0303194].

[110] S. Ashok and M. R. Douglas, "Counting flux vacua," JHEP **0401** (2004) 060 [hep-th/0307049].

[111] C. Vafa, "Evidence for F theory," Nucl. Phys. B **469** (1996) 403 [hep-th/9602022].

[112] J. J. Heckman, "Particle Physics Implications of F-theory," Ann. Rev. Nucl. Part. Sci. **60** (2010) 237 [arXiv:1001.0577 [hep-th]].

[113] T. Weigand, "Lectures on F-theory compactifications and model building," Class. Quant. Grav. **27** (2010) 214004 [arXiv:1009.3497 [hep-th]].

[114] A. Klemm, B. Lian, S. S. Roan and S. T. Yau, "Calabi-Yau fourfolds for M theory and F theory compactifications," Nucl. Phys. B **518** (1998) 515 [hep-th/9701023].

[115] W. Taylor and Y. N. Wang, "The F-theory geometry with most flux vacua," JHEP **1512** (2015) 164 [arXiv:1511.03209 [hep-th]].

[116] B. Freivogel, "Making predictions in the multiverse," Class. Quant. Grav. **28** (2011) 204007 [arXiv:1105.0244 [hep-th]].

[117] G. Obied, H. Ooguri, L. Spodyneiko and C. Vafa, "De Sitter Space and the Swampland," arXiv:1806.08362 [hep-th];
H. Ooguri, E. Palti, G. Shiu and C. Vafa, "Distance and de Sitter Conjectures on the Swampland," Phys. Lett. B **788** (2019) 180 [arXiv:1810.05506 [hep-th]].

[118] A. D. Linde and A. Mezhlumian, "Stationary universe," Phys. Lett. B **307** (1993) 25 [gr-qc/9304015].

[119] A. D. Linde, D. A. Linde and A. Mezhlumian, "From the Big Bang theory to the theory of a stationary universe," Phys. Rev. D **49** (1994) 1783 [gr-qc/9306035].

[120] M. Kleban, "Cosmic Bubble Collisions," Class. Quant. Grav. **28** (2011) 204008 [arXiv:1107.2593].

[121] A. Vilenkin, "A Measure of the multiverse," J. Phys. A **40** (2007) 6777 [hep-th/0609193].

[122] A. N. Schellekens, "Life at the Interface of Particle Physics and String Theory," Rev. Mod. Phys. **85** (2013) no.4, 1491 [arXiv:1306.5083 [hep-ph]].

[123] J. Garriga, D. Schwartz-Perlov, A. Vilenkin and S. Winitzki, "Probabilities in the inflationary multiverse," JCAP **0601** (2006) 017 [hep-th/0509184].

[124] J. Carifio, W. J. Cunningham, J. Halverson, D. Krioukov, C. Long and B. D. Nelson, "Vacuum Selection from Cosmology on Networks of String Geometries," Phys. Rev. Lett. **121** (2018) no.10, 101602 [arXiv:1711.06685 [hep-th]].

[125] D. Harlow, S. H. Shenker, D. Stanford and L. Susskind, "Tree-like structure of eternal inflation: A solvable model," Phys. Rev. D **85** (2012) 063516 [arXiv:1110.0496 [hep-th]].

[126] Y. Nomura, "Physical Theories, Eternal Inflation, and Quantum Universe," JHEP **1111** (2011) 063 [arXiv:1104.2324 [hep-th]].

[127] A. Vilenkin, "A quantum measure of the multiverse," JCAP **1405** (2014) 005 [arXiv:1312.0682].

[128] R. Bousso and L. Susskind, "The Multiverse Interpretation of Quantum Mechanics," Phys. Rev. D **85** (2012) 045007 [arXiv:1105.3796 [hep-th]].

[129] J. Hartle and T. Hertog, "One Bubble to Rule Them All," Phys. Rev. D **95** (2017) no.12, 123502 [arXiv:1604.03580 [hep-th]].

[130] F. J. Dyson, "Time without end: Physics and biology in an open universe," Rev. Mod. Phys. **51** (1979) 447.

[131] A. Guth and V. Vanchurin, unpublished.

[132] R. Bousso, B. Freivogel, S. Leichenauer and V. Rosenhaus, "Eternal inflation predicts that time will end," Phys. Rev. D **83** (2011) 023525 [arXiv:1009.4698 [hep-th]].