LPTENS-02/62
IHES/P/02/91
FIAN/TD-16/02
ITEP/TH-61/02

STRING THEORY OR FIELD THEORY? [1]

A.Marshakov

Laboratoire de Physique Théorique de l' École Normale Supérieure,[2]
24 rue Lhomond, Paris, France,
Institut des Hautes Études Scientifiques,
Bures-sur-Yvette, France

and

Theory Department, P.N.Lebedev Physics Institute,
Institute of Theoretical and Experimental Physics
Moscow, Russia [3]

e-mail: mars@lpi.ru, mars@gate.itep.ru, andrei@ihes.fr

The status of string theory is reviewed, and major recent developments - especially those in going beyond perturbation theory in the string theory and quantum field theory frameworks - are discussed. This analysis helps better understand the role and place of string theory in the modern picture of the physical world. Even though quantum field theory describes a wide range of experimental phenomena, it is emphasized that there are some insurmountable problems inherent in it - notably the impossibility to formulate the quantum theory of gravity on its basis - which prevent it from being a fundamental physical theory of the world of microscopic distances. It is this task, the creation of such a theory, which string theory, currently far from completion, is expected to solve. In spite of its somewhat vague current form, string theory has already led to a number of serious results and greatly contributed to progress in the understanding of quantum field theory. It is these developments which are our concern in this review.

[1]Based on contribution to *Physics Uspekhi*, **172**, 2002
[2]Unité mixte de Recherche 8549 du Centre National de la Recherche Scientifique et de l'Ecole Normale Supérieure
[3]Permanent address

Contents

1 Introduction

The 20th century may be considered as a century of success (*uspekhi*) for physics. Absolutely new physical ideas about the world which surrounds us have greatly affected every human being and indeed the whole of mankind, especially those people in power. This is shown by the wide spread use of radio and television, man going into space, and – perhaps chiefly – by explosions of atomic and hydrogen bombs. Thus, originally found "with pen and paper" electrodynamics, the theories of relativity and quantum mechanics have completely proved their worth.

Probing further into the "deep secrets of the world" in an attempt to understand the very small – subatomic and subnuclear – structure of our world, has not proved straightforward. The absence of an experimental base, or at the very least, big problems with experiments directed to check any statement about energies more than 100 GeV, has led to the situation where theoretical physics has relied more and more upon its "internal beauty". In other words, it develops, in a fashion similar to mathematics, mostly based on its own logic. As a result of such developments, one had by the end of the 20th century a situation quite rare for physics before. This search for "internal harmony" among theoretical physicists distanced them quite far from the desires of experimentalists, at least in the field of elementary particle physics. The so called Standard Model (unifying theory of electromagnetic and weak interaction based on the Weinberg-Salam model and chromodynamics) appears to be almost completely satisfactory [1] from the point of view of all known experiments. Already for about thirty years theoreticians look for a "nice fundamental" theory, which reproduce the Standard Model at large distances or energies of the order of W-boson mass (roughly, the same 100 GeV). Despite obvious weaknesses of the arguments about "beauty" as a foundation for theoretical physics, the majority of interested people including myself can say that the Standard Model is not satisfactory only from the point of view of this principle. Moreover, already within the framework of the Standard Model a few ideas were used (spontaneous breaking of the gauge symmetry or the Higgs effect), which are not yet confirmed by experiment but were rather chosen among all possible options only due to their beauty and simplicity. In this way, the Standard Model W-bosons become massive due to interaction with the condensating scalar field, in complete analogy with the Landau-Ginzburg mechanism in the condensed matter physics, though the excitations of this scalar field have never been seen in nature.

Hence, in this review we will try to discuss the theory, which cannot be verified by experimental particle physics. In this sense this hypothetical theory is somehow more close to gravity than to elementary particle physics, where after the appearance of General Relativity "internal beauty" plays the role of the main physical principle. In the theory of gravity, which is responsible mainly for the physics of the *macro*world, the separation from experiment (or, better to say, lack of experimental base for fixing the parameters of the theory) has always allowed the possibility of using some extra purely "internal" theoretical principles. It turns out, that such a situation permeates also more and more into the physics of the *micro*world.

A natural requirement to such a hypothetical theory would be an explanation of "everything" including gravity (which is definitely beyond the Standard Model), i.e. the formulation of all four interactions – electromagnetic, strong, weak and gravitational – starting from some unique principles. This review contains an attempt to formulate these general principles and to demonstrate that they could lead to some progress not only in understanding of quantum theory of gravity, but also to some absolutely new perspective on the well-known problems in gauge theories, being the base of the Standard Model. It is certainly clear that there cannot be any "uniqueness theorem" for such hypothetical fundamental physical theory and therefore everything to be said below, especially without direct experimental confirmation, can be considered as a "pure fantasy". We will try to show nevertheless that it is this particular variant of such a "fantasy" which is based on relatively simple and clear physical principles (though not always clearly formulated), which become especially attractive when taking into account that all alternative attempts to achieve any progress at least in qualitative understanding of microworld physics, up to now have been totally unsatisfactory.

Mostly for historical reasons the fundamental theory at small distances of the Planck scale $\sqrt{\frac{\gamma_{\rm N}}{\hbar c}} \sim 10^{-33}$ cm ($\hbar$ and c denote the Planck constant and vacuum speed of light [2], while $\gamma_{\rm N}$ is the Newton gravitational constant), where it is necessary to take into account effects of quantum gravity or, stated alternatively, gravity becomes comparable with the other interactions, is called *String Theory* [3]. This name can be considered not ideal and

[1] Precise checks of the predictions of the Standard Model have not found any contradictions between the theory and experiment, coming out of three standard deviations, what is quite satisfactory since, as L.B.Okun reminded me, Landau and Fermi suggested always multiply the errors of experiment by π. The latest data can be found in the report by M.Grünewald (Talk at LEP Physics Jamboree, CERN, July 10, 2001) available at http://www.cern/ch/LEPEWWG.

[2] In what follows, if not specially noticed, these constants are formally put equal to unity, i.e. in relativistic physics of microworld velocities will be measured in units of speed of light c, while actions in units of the Planck constant $\hbar$.

[3] There exists already vast literature on string theory including few books by people who "founded" the string theory (see

other suggestions for different names show up from time to time (say, M-theory etc). In what follows we will use the "traditional" name, since though being not complete or exact term, it "catches" in the best way one of the main principles of this theory – a natural "geometric" regularization of small distances by introducing the extended objects of non-zero length (typically of the Planck scale). The appearance of strings in the role of such extended objects immediately leads to the theory containing massless gauge bosons and gravitons (whose consistence though has yet to be proven).

Let us point out separately that the widely used (especially in popular literature) word "superstrings", seems to be much more unacceptable because, first, it literally corresponds only to the narrow class of string models and, second, it mixes two absolutely different and mutually independent physical ideas. It couples the concept of strings proper with the very different idea of supersymmetry (or symmetry between bosons and fermions). As we see below, the role of such symmetry is especially important in quantum field theory, where supersymmetry allows even to extend the horizon of applications of the Standard Model. In contrast to typically field-theoretical role of supersymmetry aimed to cancellation of the ultraviolet divergencies, for string theory of main importance is that fundamental theory at small distances is *not* a local quantum field theory, and this is already encoded in its name.

Let us specially stop at this point. On one hand, string theory does not contradict to the existence of quantum field theory as a reasonable effective theory at energies much less than Planckian (10^{19} GeV), which naturally describes the physical processes at weak coupling. Within its range of validity, quantum field theory automatically takes into account the contribution of anti-particles and proposes the values for the amplitudes and cross-sections which are in rather nice agreement with experiment. Moreover (and this will be discussed below in detail), when studying the processes where the contribution of gravity is inessential or at energies much less than Planckian energy, string theory often reduces to quantum field theory – to the theory of gauge vector fields. It is exactly in this sense the field theory is often called an *effective* theory for strings at large distances. Roughly speaking, field theory arises in the low-energy limit of string theory, similar to how non-relativistic limit of the field theory gives rise to quantum mechanics, which in its turn as $\hbar \to 0$ reduces to classical mechanics.

On the other hand, one should immediately notice that historically the step towards the string theory from quantum field theory is nothing else but change of the *paradigm*, and within the frames of new paradigm quantum field theory can no longer pretend to the role of fundamental physical theory. Below we are going to discuss this point applying rather simple physical principles, which lead to an understanding that any attempts to construct theory of quantum gravity in the framework of quantum field theory are almost absurd.

However, here one should definitely and honestly point out that the situation within string theory itself is far from being perfect. Pretending to be the fundamental theory of microworld and unifying theory of all interactions, string theory has not only been formulated in closed form, but even does not have any well-studied "sample example", demonstrating more or less all its basic ingredients, like in the case of simple models of quantum mechanics (a harmonic oscillator or an atom of hydrogen) or quantum field theory (say, scalar field theories with ϕ^3- or ϕ^4-potentials, or quantum electrodynamics). In fact, at present only some "pieces" of string theory, rather chaotically placed among other "pieces", are available to be investigated and partially formalized. Nevertheless, during recent years some definite progress has been detected (and is still taken place!) in the area of string theory, which certainly distinguishes it among other, practically dead-end directions.

The main purpose of this review is to discuss basic *physical* principles forming the base of string theory and try to demonstrate their attractive features, reviewing some (in particular recent) achievements in this sphere. Notice immediately, that these achievements are not at all obvious to everybody and do not explain (yet?!) observable physical phenomena. It seems nevertheless to be very important that only in the framework of string theory at least the possibility to *raise* several new questions of principal importance arose. One of the most well-known of them is the problem of the space-time dimension, supposed to be solved *dynamically* instead of usual fixing of the dimension "by hand". This approach is totally new in comparison with traditional point of view accepted in quantum field theory, where space-time belongs to a few initial basic ingredients.

The dynamical nature of space-time is a direct consequence of *definition* of string theory already at perturbative level by the Polyakov path integral where the sum over all physically different configurations is represented by the sum over all geometries on two-dimensional string world-sheets. The arising "geometrization" of string theory already at the perturbative level also plays an essential role in the attempts to go beyond the perturbation theory. Recent most striking achievements are indeed related to the ideas to identify parameters of physical theory (masses, condensates, coupling constants) with the parameters or moduli of certain (complex) manifolds arising as a "compact parts" of the full space-time, dynamically chosen by string theory.

[1, 2, 3]) lots of review, including the reviews in Physics Uspekhi [9, 10] etc. However this branch of science is still developing so rapidly that re-understanding even of the basic notions and concepts happens quite often. Say, relatively new reviews [14]-[21] are quite useful my point of view, though this is certainly the list far from being complete.

To finish this introduction let us also point out that the specific situation around string theory, quite untypical for physics, also leads to a large amount of "social" problems, which are rather interesting in themselves but their discussion goes beyond the scope of this review. For example, string theory very often (and at least from my point of view very unfair) is claimed to be "pure mathematics" in contrast to many other, more traditional spheres of activity in theoretical physics considered to be "physics by definition". In particular, many physicists got used to the more traditional paradigm of quantum field theory and call *all* problems of string theory "mathematics" only because they arose in this particular context, while any technical problem of the formalism of quantum field theory is considered as "physics".

It is certainly true that string theory as any other interesting sphere in theoretical physics raised lots of new mathematical problems and requires the application of branches of mathematics previously not widely used in physics, moreover certain problems of string theory are playing the role of "locomotive" for some directions of research in mathematics. However, it seems to be completely wrong to stress this particular aspect of the new theory and in what follows we will try to discuss mostly simple and natural *physical* aspects of string theory.

Another social effect which is quite often (and again unfairy) associated only with string theory is the widely spread invasion of "marketing" principles into the modern science. Caused by purely social problems, continuous advertisement of the string theory as a theory which has *already* solved all possible problems of natural science (especially on the background of absence of any strict arguments supporting this point of view) does great harm to anybody willing to understand seriously this interesting direction in modern physics. Together with the lack of relations with experiment, existing for more traditional spheres of theoretical physics, the wide advertising of string theory brought only negative attention to this field of science especially among quite conservative physicists. However, it is also necessary to stress that the development of string theory in present conditions would be simply impossible without bright and striking new ideas (see, for example, [1]), which only partially, and mostly many years after they had been pronounced, were turned into the frames of more or less strict formulations. It is rather natural to get a lot of "garbage" along this way and one of the main difficulties is the opportunity to be killed by huge stream of various literature which often does not contain any useful information. Without pretending to objectivity, especially in such a delicate question, I certainly understand that the reference list to this review contains only very restricted fraction of existing literature, and the choice of these particular references was mostly determined by (sometimes accidental) my personal knowledge.

Content of the review. We start in sect. 2 with discussion of the Standard Model of gauge interactions (electromagnetic, weak and strong) of elementary particles and (classical) theory of gravity – General Relativity. The main aim of this discussion is to fix once more the status of quantum field theory as absolutely satisfactory and verified experimentally model of observable interactions of elementary particles, which however runs into serious difficulties in the strong coupling regime and, mainly, which is absolutely useless as a theory of quantum gravity.

In sect. 3, we will try to formulate the main principles of string theory, coming mostly from geometric formulation of string perturbation theory in terms of the Polyakov path integral. The main message of this section is that it is two-dimensional geometry – the basic point of the Polyakov formulation – which is responsible for new string approach to the dynamical nature of space-time and here is principle difference between string theory and standard quantum field theory. We will also discuss supersymmetry as an origin for appearance of the fermions and the Fradkin-Tseytlin effective actions, being the most convenient "bridge" between string theory and effective quantum field theories.

Sect. 4 is devoted to recent attempts in string theory to go beyond the perturbative regime. The main purpose of this section is to explain the basic ideas of these attempts: the idea of duality between the theories at strong and weak coupling and the classical extended objects appearing necessarily in non-perturbative string theory. As an illustration of the progress in studying the non-perturbative effects being an outcome of applying new stringy methods, we will discuss the Seiberg-Witten theory which allows, in particular, to make a new step in understanding of the mechanism of confinement.

Sect. 5 is totally devoted to one of the most interesting new problems in string theory – an attempt of dual description of the non-Abelian gauge theories at strong coupling in terms of gravity (or theory of closed strings). Finally, in sect. 6, we review a few other modern directions coming out of string theory, this section being written for the most advanced reader (the same is true for the sect. 4.6). Paragraphs of the text containing technical issues and therefore being more difficult for understanding, are typed with a smaller font.

2 Physics of Elementary Particles. Gauge Theories and Gravity

There have been no essential changes in elementary particle theory during the last decades. Still two main problems are at the center of interest, these are confinement (or keeping of quarks locked inside the hadrons) and the quantum theory of gravity [4], while all the rest can be almost completely explained in the framework of the Standard Model. Mostly probable, the solution of these two problems is impossible without progress in understanding of the properties of gauge theory and general relativity at strong coupling, i.e. exactly where the standard field-theoretical methods being the basic ones for the Weinberg-Salam model of electroweak interactions and quantum chromodynamics (QCD) at high energies become useless.

The Standard Model in its main features can be considered as a non-Abelian gauge theory with the gauge group $SU(2) \times U(1) \times SU(3)$ (the last factor corresponds to the "color" or strong interaction) and matter fields of "three generations" [5] (see also, e.g. [6]). The computations are performed using the technique of the gauge field theory [4] at small coupling constants – i.e. by perturbation theory, and the results of such computations are nicely consistent with experiment (see footnote 1 [5]). From pure theoretical or kind of aesthetic point of view the Standard Model is a little bit "ugly" due to presence of "external" parameters – such as the Weinberg angle, as well as due to absence of completeness in some questions like spontaneous symmetry breaking or the Higgs effect, which is responsible for masses of non-Abelian W- and Z-bosons. Nevertheless, the Standard Model is an absolutely consistent quantum field theory. It is a renormalizable quantum field theory, which was already marked by the corresponding Nobel Prize in physics [37].

If speaking about gravity, its "observable part" is still negligible in the sense of the possible influence or this or that choice of the theory of quantum gravity. At least to my knowledge by now there is no direct experimental evidence of the existence of gravitons as well as any clear and unambiguous data concerning the problems of dark matter and cosmological constant (see, for example, [36]). All experts agree only that dark matter seems to exist and the cosmological constant looks like being nonvanishing. Despite of growing precision of experimental methods in astrophysics, the existing data are too scarce in order to put at least some framework onto the set of existing theoretical models. Moreover, the very idea of applicability of present physical theories to the model of Universe as a whole seems to be rather "voluntaristic", while the attempts to formulate the model of Universe in terms of microworld physics, i.e. in the language of quantum mechanics or quantum field theory do not have any real physical background and can be considered almost absurd. Thus, when discussing the problems of quantum gravity one has to use only pure theoretical and aesthetic criteria.

2.1 Gauge Field Theories

Gauge theories or theories of massless vector fields describe all interactions except for gravity. The theory of gauge fields or the Yang-Mills fields can be formulated without even using stringy principles and can be considered as a closed physical theory within some range of energies. Nevertheless the viewpoint onto the theory of gauge fields as being "derivative" from string theory leads to its much deeper understanding and already brought us to new interesting results.

The progress achieved in gauge theories, especially in their supersymetric versions has allowed many people to say that gauge theory, or even any quantum field theory can be treated beyond the level of perturbative expansion. However, and we want to stress this point, when such words are pronounced it is usually implied that something *extra* should be added to the standard definition of the quantun field theory, base on particular well-known manipulations with field-theoretical Lagrangian. In other words, this implies some *new* definition of a quantum field theory (rather different from a standard one) which is even more close to what we call here by string theory. Despite this seems to be only a terminological difference, using of the old expression "field theory" is not quite adequate in this case, because the "new definition" of field theory in practice leads to change of the paradigm, since the new effects cannot be obtained on the level of formal manipulations with the Lagrangian.

In gauge theories matter interacts due to exchange by massless vector fields. In case of electrodynamics or Abelian theory the gauge group is $U(1)$, i.e. it contains the only vector field is $A_\mu(x)$ (photons), if the theory is non-Abelian (or equivalently the Yang-Mills gauge theory [4]) the fields can be conveniently represented by matrices from the Lie algebra of the corresponding gauge group $\mathbf{A}_\mu(x) \equiv \|A_\mu^{ij}\|$ (gluons), in the $SU(N)$ case, for example, by the $N \times N$ (anti)Hermitean traceless matrices. The minimal interaction is introduced by the "long" derivative

$$\partial_\mu \to \mathbf{D}_\mu = \partial_\mu + \mathbf{A}_\mu \tag{2.1}$$

[4] More strictly these are the problems of elementary particle physics "in a wide sense". From a more "narrow" point of view one may in principle doubt in existence of the problem of quantum gravity.

[5] Apart from neutrino oscillations (see, for example, [35]).

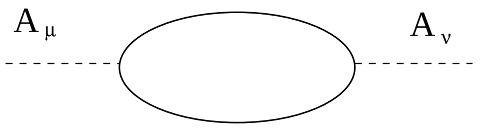

Figure 1: *One-loop diagram arising in the calculation of corrections to the effective charge. In electrodynamics, as follows from Ward identities, the computation may be restricted to only this diagram*

or $D^{ij}_\mu = \partial_\mu \delta^{ij} + A^{ij}_\mu$, if the gauge field interacts with matter from the representation of the gauge group whose elements are labeled by index i. The gauge-invariant Lagrangian of the Yang-Mills fields has the form

$$\mathcal{L}_{\rm YM} = \frac{1}{2g^2}{\rm Tr}\mathbf{F}^2_{\mu\nu} \tag{2.2}$$

where

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + [\mathbf{A}_\mu, \mathbf{A}_\nu] \tag{2.3}$$

In the case of electrodynamics matrix-valued fields turn into numbers and therefore the formula (2.3) does not contain commutators (leading to the self-interaction in (2.2)) and one may not write the trace Tr over the matrices.

For the Standard Model the gauge group is $SU(3) \times SU(2) \times U(1)$ and one should add to Lagrangian (2.2) the Lagrangian of matter fields (electrons, quarks, etc) with the "long" derivative (2.1). After that one can perform the standard field-theoretical computations developing the perturbation theory in coupling constant g. Such a theory will no more be fundamental at the level of field-theoretical perturbation theories, since it contains the Abelian factor $U(1)$ with coupling constant growing at small distances, while the theory with "controlled" behavior at small distances "should be" non-Abelian. In what follows we will restrict ourselves to the compact (for the integrality of charges!) non-Abelian $SU(N)$ groups, considering all other gauge groups as "pure exotic".

The reason for the "non fundamental" nature of the Abelian theories is famous "zero-charge" or "Moscow zero" in electrodynamics. In quantum field theory parlance this means the growth of charge at small distances. The physical origin of such behaivior comes from the screening of charge by virtual electron-positron pairs, while the gauge $U(1)$ fields themselves are not charged. Technically this means that one-loop corrections (the simplest diagram for the computation of this effect, say, in electrodynamics is depicted in fig. 1) lead to the following dependence of effective charge on the energy scale μ

$$\frac{dg}{d\log\mu} \equiv \beta(g) = b_0 g^3 + \ldots \tag{2.4}$$

where the coefficient

$$b_0 \propto N_F - N_V \tag{2.5}$$

is the *difference* of the contributions N_F of matter fields and N_V of the gauge fields themselves, propagating along the loop at the diagram in fig. 1. In electrodynamics the self-interaction of photons is absent, hence $N_V = 0$, and the coefficient in formula (2.5) is positive. This means the growth of charge with μ, or approaching small distances and as a consequence electrodynamics at small distances is not well-defined, i.e. cannot be a fundamental theory. Simultaneously electrodynamics continues to be nice effective theory at *large* distances, where $g_{QED} \equiv e$ is small [6].

The situation changes drastically for the case of *non* Abelian gauge theories where extra *anti* screening of charges by *charged* (in color) gauged fields exists so that $N_V \neq 0$ due to self-interaction of gluons. This leads to the possibility of "asymptotic freedom" [65], when interaction becomes weak at small distances for $N_F < N_V$. The difference is demonstrated in fig. 2, where the difference between zero-charge and asymptotically free theories can be clearly seen.

A natural way out from such situation is to consider electrodynamics as a "part" of some non-Abelian theory from which is "splits" at some scale where non-Abelian symmetry is violated. In such a case the non-Abelian gauge theory (especially in supersymmetric case) can be considered as "fundamental", at least in some energy range where the effects of gravity have not yet become necessarily taken into account. From this perspective *renormalizability* of gauge theories has a quite simple meaning – the Lagrangian (2.2) is useful for the description of physics in rather large energy range if for the coupling g one would substitute its corresponding effective value at given energy. With this substitution the general form of the Lagrangian remains intact (and it does not require additional terms when passing from one energy to another).

[6]This is true in the elementary particle physics, but not in condensed matter theory, where instead of $\frac{e^2}{mc} \sim \frac{1}{137}$ the parameter of perturbative expansion is $\frac{e^2}{mv_F} \sim 1$.

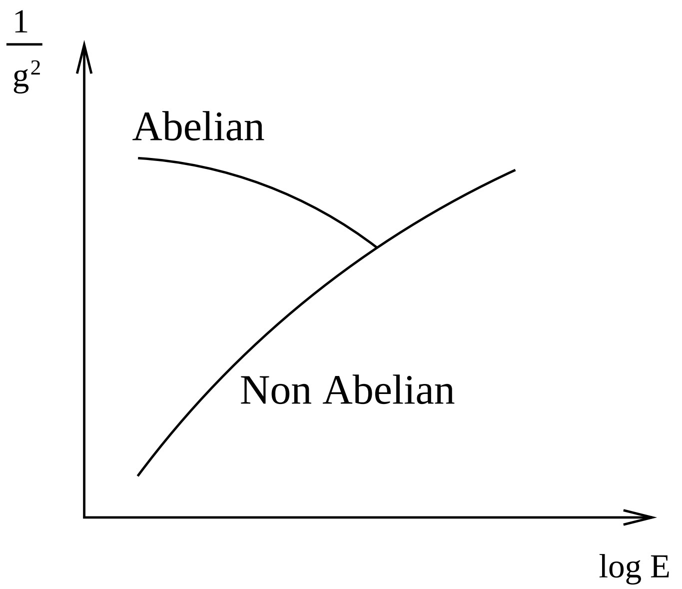

Figure 2: *Dependence of effective couplings upon energy in Abelian and non-Abelian gauge theories. The top curve corresponds to "zero charge" while the bottom curve corresponds to asymptotic freedom.*

2.2 Spontaneous Breaking of Gauge Symmetry

Let us now discuss how at some energy scale the gauge group can (partially) turn into Abelian. In the most natural way it happens if the theory contains the scalar fields in the adjoint representation of the gauge group, for example, as a consequence of supersymmetry. Suppose the scalar potential has minima such that condensates or vacuum expectation values of scalars do not vanish. For the field in adjoint representation of the gauge group $SU(N)$ this means that vacuum values $\boldsymbol{\phi}$ may be chosen in diagonal form

$$\boldsymbol{\phi} = \left.\begin{pmatrix} \phi_1 & & & \\ & \phi_2 & & \\ & & \dots & \\ & & & \phi_N \end{pmatrix}\right|_{\mathrm{Tr}\boldsymbol{\phi}=\sum \phi_j=0} \tag{2.6}$$

using gauge invariance. For the convenient choice of gauge-invariant quantities one may take parameters like $\mathrm{Tr}\boldsymbol{\phi}^k$ or their "generating functions"

$$P_N(\lambda) = \det(\lambda - \boldsymbol{\phi}) = \prod_{i=1}^{N}(\lambda - \phi_i) \tag{2.7}$$

The total number of algebraically independent parameters $\{\phi_i\}$ is equal to the rank of the group, in the mostly well-known case this is $\mathrm{rank}[SU(N)] = N-1$. It is customary to say that these parameters are co-ordinates in the parameter space or moduli space of gauge theory. Due to the Higgs effect the off-diagonal part of the matrix of gauge field $\mathbf{A}_\mu$ for $\boldsymbol{\phi} \neq 0$ becomes massive, since the interaction

$$[\boldsymbol{\phi}, \mathbf{A}_\mu]_{ij} = (\phi_i - \phi_j)\mathbf{A}_\mu^{ij} \tag{2.8}$$

literally turns into the mass terms

$$\sum(\phi_i - \phi_j)^2 \left(\mathbf{A}_\mu^{ij}\right)^2 = \sum(m_{\mathrm{W}}^{ij})^2 \left(\mathbf{A}_\mu^{ij}\right)^2 \tag{2.9}$$

in the Lagrangian. At the same time the diagonal part, as follows from (2.9), remains massless, i.e. the gauge group $G = SU(N)$ is broken by Higgs mechanism to $U(1)^{\mathrm{rank}G} = U(1)^{N-1}$ [7].

Thus, in the generic situation at the scale $\boldsymbol{\phi}$ (the scalar field has a dimension of mass) non-Abelian gauge group is broken down to Abelian which in the simplest $SU(2)$ case is exactly that of electrodynamics. In what follows, even in general $U(1)^{N-1}$ case we would call such an Abelian theory (generalized) electrodynamics and refer to the corresponding charges as electric charges.

[7] In the situation of "general position", i.e. when $\phi_i \neq \phi_j$ for $i \neq j$. If the eigenvalues (2.6) partially coincide, the broken group still contains non-Abelian factor $SU(K)$ with $K < N$.

2.3 Nonperturbative Effects: Instantons and Monopoles

In contrast to electrodynamics the non-Abelian gauge theories are essentially nonlinear since the Lagrangian (2.2) contains cubic and quartic terms in the Yang-Mills fields. It means that equations of motion are nonlinear even without the matter fields. Nonlinear equations typically do have lots of nontrivial solutions, related in the case of non-Abelian gauge theories to nontrivial topological properties of the gauge groups.

Do these solutions affect elementary particle physics? The exact answer to this question is still only hypothetical, but from general arguments it is clear that the influence can be essential in the strong coupling regime. Indeed, from general properties of quantum theory we know that the main contribution of a classical trajectory to quantum amplitude (the Feynman path integral) is nothing but $\exp(-S/\hbar)$, where S is the classical action on given configuration. For the theory of non-Abelian gauge fields the corresponding action, or Lagrangian (2.2), integrated over space-time, will give rise to the contributions of the form $\exp\left(-\frac{\mathrm{const}}{g^2}\right)$, which are exponentially suppressed at weak coupling. However, by the same logic it is quite possible, that the same contribution would be much more essential at strong coupling, i.e. exactly there, where the main and unclear yet phenomena are "hidden". Hence, the classical solutions look like being very important for studying the strong-coupled phase.

At present among all classical solutions in non-Abelian gauge theories the most essential role belongs to *instantons* or pseudoparticles [69, 70, 32]. By instanton one usually means the configuration of fields "localized" in four-dimensional Euclidean space, which satisfies the (anti) self-duality equations

$$
\begin{aligned}
&\mathbf{F} = \pm {}^*\mathbf{F} \qquad \text{Ë«Ë} \\
&\mathbf{F}_{\mu\nu} = \pm\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\mathbf{F}_{\lambda\rho} \equiv \pm\widetilde{\mathbf{F}}_{\mu\nu}
\end{aligned}
\tag{2.10}
$$

($\mu,\nu = 1,\ldots,4$). Any solution to the self-duality equations (2.10) is automatically a solution to the Yang-Mills equations of motion $\mathbf{D}_\mu\mathbf{F}_{\mu\nu} = 0$ (the opposite is incorrect!) due to the Bianchi *identities* $\mathbf{D}_\mu\widetilde{\mathbf{F}}_{\mu\nu} = \epsilon_{\mu\nu\lambda\rho}\mathbf{D}_\nu\mathbf{F}_{\lambda\rho} \equiv 0$ (i.e. relations, true for any fields). For the instantons

$$
S = \frac{1}{2g^2}\int_{d^4x}\mathrm{Tr}\mathbf{F}_{\mu\nu}^2 = \frac{1}{2g^2}\int_{d^4x}\mathrm{Tr}\mathbf{F}_{\mu\nu}\widetilde{\mathbf{F}}_{\mu\nu} = \frac{8\pi^2 n}{g^2}
\tag{2.11}
$$

where n is the topological charge, counting how many times the three-dimensional sphere of large radius in four-dimensional space-time "winds" around the compact gauge group (in fact around its $SU(2)$ subgroup). We will see below that in certain important examples the nonperturbative configurations in some sense are "exhausted" by instanton configurations.

The simplest one-instanton solution [70] to the self-duality equations (2.10) has the "bell-shaped" form

$$
\begin{aligned}
&\mathbf{A}_\mu \propto \boldsymbol{\eta}_{\mu\nu}\frac{x_\nu}{x^2+\rho^2} \\
&\mathbf{F}_{\mu\nu} \propto \boldsymbol{\eta}_{\mu\nu}\frac{\rho^2}{(x^2+\rho^2)^2}
\end{aligned}
\tag{2.12}
$$

in four-dimensional space-time with the center, chosen in (2.12), to be at the point $x_0 = 0$. In eq. (2.12) we have introduced $\boldsymbol{\eta}_{\mu\nu}$ – the 't Hooft **C**-number matrices (see, for example [32]). Solution (2.12) corresponds to the topological charge of the instanton $n = 1$.

Another important nonperturbative effect is the monopole or a particle with magnetic charge. In the Abelian theory monopoles can arise only as external sources, but in the framework of non-Abelian theory they can be identified with certain configurations of extra (scalar or Higgs) fields [67]. The simplest monopole configuration arises as a result of reduction of the self-duality equation (2.10), when fields do not depend on time and $\mathbf{A}_0 = \mathbf{\Phi}$ is considered as an extra scalar. Under such reduction the self-duality equations (2.10) turn into the Bogomolny equations

$$
\mathbf{D}_i\mathbf{\Phi} = \frac{1}{2}\,\epsilon_{ijk}\mathbf{F}_{jk}
\tag{2.13}
$$

($i,j = 1,\ldots,3$). As in the instanton case the topological configuration of monopoles is nontrivial – they cannot be obtained by continuous deformation of configurations with trivial (vanishing) fields. The obstacle is topological charge. The monopole masses are similar to the actions of instanton configurations. For the so called BPS-monopoles [71], being exactly the solutions to equations (2.13), the masses are equal to

$$
m_{\mathrm{mon}}^{ij} = \frac{4\pi}{g^2}m_{\mathrm{W}}^{ij} = \frac{4\pi}{g^2}\left(\phi_i - \phi_j\right)
\tag{2.14}
$$

It follows from this formula that at weak coupling the monopoles are very *heavy* particles. However, the situation can again change after passing to the strong coupling area, though the formula (2.14) is literally incorrect. However, at strong coupling the monopoles might become even more *light* than ordinary, i.e. electrically charged particles. In such circumstances the condensation of light monopoles can bring us to confinement of electric charges similar to the Meissner effect in superconductivity.

Thus, the nonperturbative effects related to nontrivial classical configurations may play an important role when describing the theory at strong coupling. On of the attendant technical problem is that these effects are usually "screened" by the perturbative corrections. In order to get a clearer picture of the nonperturbative effects one should pass to supersymmetric theories [63, 76, 66] (see also the papers [48, 49, 75], the books [7, 2] and reviews [42, 31, 40, 39]).

2.4 Supersymmetric Gauge Theories

The main distinguishing feature of the supersymmetric theories is that they contain an equal number of bosonic and fermionic excitations. Therefore, due to the different signs of the bosonic and fermionic contributions into loop diagrams one gets essential cancellation of divergences. This effect is easily seen, say, directly in the formula (2.5), if one puts N_F to be the contribution of fermionic loops, while N_V – the contribution of bosonic loops.

Adding to the corresponding Lagrangians the superpartners of the vector and matter fields one may consider non-Abelian gauge theories as quite satisfactory for the description of all interactions (except for gravity) in some vast range of energies. Renormalizability still means that theory is described by (supersymmetric) Lagrangian of the Yang-Mills fields with matter terms added in some range of scales and the *only* thing to be added to such Lagrangian is prescription how the coupling $g = g(\mu)$ depends on the scale μ. This is governed by the renormalization group equation (2.4), which looks much simpler in supersymmetric theories due to cancellation of loop corrections in perturbation theory.

One of the main "phenomenological" problems of supersymmetric gauge theories [8] is the presence of scalar fields in their spectra. The scalar fields are necessary superpartners for the matter fermions and even for the Yang-Mills fields in the case of extended supersymmetry, i.e. when each field has more than a single superpartner. Due to supersymmetry the excitations of the scalar fields should have the same masses as the excitations of fermions (and vector fields) which totally contradicts to the observable spectrum in nature. It means that in our world supersymmetry is broken at least at some scale and the dynamical derivation of such a scale is one of the main problems of the theory. However, if we believe that this problems will be solved, beyond this scale (at small distances) the supersymmetric theory is a good object for study since it is not so "polluted" by loop corrections.

In contrast to nonvanishing vacuum expectation values of the other fields the scalar condensates $\langle\phi_A\rangle \neq 0$ do not violate the space-time symmetry. Then in low-energy effective theory all parameters of the effective Lagrangian (masses, couplings) become in general nontrivial functions of these condensates. As we already mentioned such functions are usually called functions on the moduli space of supersymmetric gauge theories. In gauge theories with extended supersymmetry (when number of supersymmetry generators in terms of the Majorana spinors is $\mathcal{N} = 2$ and higher) one cannot write down potential energy for Abelian fields not violating supersymmetry. In non-Abelian theories the only choice for such a potential term, not violating extended supersymmetry, is to take the sum of commutators of the matrix-valued fields $\{\phi_A\}$ of the form $\sum_{A<B} \mathrm{Tr}[\phi_A, \phi_B]^2$. In theories with such potential energy only the light Abelian fields "survive" at large distances, i.e. one gets electrodynamics (see (2.9)) together with massless scalars or moduli – the fields whose vacuum values can be arbitrary. Hence, in gauge theories with extended supersymmetry there exists an infinite number (parametric family!) of vacua and the problem of the theory is to find the spectrum and effective couplings of the low-energy effective theory as functions of the vacuum condensates. An important circumstance is that supersymmetry imposes extra requirements on the space of condensates, in particular this space should be complex (and sometimes moreover Kähler, special Kähler or hyper-Kähler) so that the class of available functions is essentially restricted. All these general arguments are applicable only in the case when supersymmetry (or any other symmetry) is the exact symmetry of quantum theory, i.e. is not violated by quantization.

In the theories with "minimal" $\mathcal{N} = 1$ supersymmetry the Abelian superpotential is generated and moduli, in general, become massive and acquire fixed vacuum expectation values. In complex co-ordinates on moduli space the superpotential is a holomorphic function $W(\phi_A)$, and vacua are defined by the equation $dW = 0$,

[8] The phenomenology of supersymmetric quantum field theories goes beyond the scope of this review (see, for example recent review in Physics Uspekhi [39]). This is a quite interesting and fashionable topic, whose only weak point is the absence of experimental confirmation of supersymmetric particles. From our point of view it is much more important that supersymmetric theories play the role of a nice "theoretical laboratory" for studying nonperturbative effects in realistic gauge theories.

since potential $V(\phi,\bar{\phi}) \propto \sum_A \left|\frac{\partial W}{\partial \phi_A}\right|^2$. The geometrical meaning of the appearance of the complex manifolds in field theory is absolutely unclear, but, as we see below, it is rather natural to consider this phenomenon as an "artefact" of string theory. It is very nontrivial that complex geometry sometimes allows one to predict the *exact* form of the low-energy effective Lagrangians which already account for the nonperturbative effects (see sect. 4.5 below).

2.5 General Relativity as Effective Theory

The discovery of instantons and other nonperturbative solutions essentially extended the behavior of the theory of strong interactions. It has been demonstrated that the elementary particle physics does not reduce to perturbation theory, whose frames in QCD are determined by high energies (the asymptotic freedom regime), where the standard formulation of the gauge field theory based on perturbation theory works quite well [4]. Nevertheless, the instantonic computations appeared to be only the next approximation in QCD far not enough to describe confinement and other effects of strong coupling. As for quantization of gravity, even supersymmetry [66, 7] as a mechanism for cancellation of divergencies does not allow any dream about the possibility of a consistent theory of quantum gravity in the framework of quantum field theory (see, for example, [42]). Despite many attempts to construct a theory of quantum gravity in the framework of quantum field theory, say, as a field theory with infinite-dimensional group of gauge symmetry, such an approach seems to be based on nothing for a few quite simple reasons. We will try to discuss these reasons in this section and will come back to them many times below when speaking about string theory.

Let us first notice that by quantum field theory, if nothing opposite is stated directly, we will understand the *local* quantum field theory, satisfying the renormalizability criterium. The local quantum field theory (with Lagrangian depending upon not higher than second derivatives) guarantees a well-defined procedure of quantization of a *free* field – an infinite system of particles and anti-particles, corresponding to the quadratic in fields part of the Lagrangian. The interaction in such a theory is introduced by terms of higher degree in the fields and in weak coupling approximation the relativistic quantum field theory nicely describes the scattering of particles. It automatically takes into account the contribution of antiparticles into the physical processes, which can be considered at present as its main achievement.

A much more delicate aspect is renormalizability – the dependence of coupling constants upon the energy scale. In a renormalizable quantum field theory an interaction can be described by a *finite* set of couplings (often even a single coupling, as in gauge theories, see sect. 2.1), whose dependence of scale is rather *weak*. In reality this "weak dependence" means logarithmic dependence of the dimensionless coupling constants, like in gauge theories or $\lambda\phi^4$-theory in four dimensions. Renormalizability means that in some wide range of energies the theory is described by a single Lagrangian – new interaction vertices should not be added and the corresponding couplings weakly depend on the scale.

In the theories with dimensional coupling constants and/or an infinite set of interaction vertices these features lose any sense. The dimension of coupling constant, more exactly "negative mass" dimension like the dimension of the Newton gravitational constant $\gamma_{\rm N} \sim 1/M^2$ in four dimensions (in D-dimensional space-time $\gamma_{\rm N}^{(D)} \sim M^{2-D}$) leads to unbounded growth of the perturbative corrections of the form

$$1 + \gamma_{\rm N}\Lambda^{D-2} + \ldots \tag{2.15}$$

when one removes the cutoff $\Lambda \to \infty$. This means that the theory at any finite scale depends on what happens at small distances. This completely contradicts the idea of renormalizability, i.e. the idea that after introducing scale-dependence of the couplings one may completely forget about small distances.

Such a concept appears to be totally acceptable for renormalizable (supersymmetric) gauge theories, but is absolutely useless for the theory of gravity. Gravity (with dimensional coupling and infinitely many interaction vertices of gravitons $h_{\mu\nu}(x) = G_{\mu\nu}(x) - \delta_{\mu\nu}$) "remembers" small distances and is *not* renormalizable field theory. The same conclusion follows from the study of "lattice" or discretized gravity (except for two-dimensional case [60, 97, 98, 99], directly related to string theory), where the continuum limit is not well-defined, in contrast to, say, lattice gauge theories.

The difference between gravity and quantum field theory is in fact far deeper. Quantum field theory computes only the "relative" but *not* "absolute" value of a physical quantity, i.e. only the difference between the value of some quantity at given scale μ and its value at some "normalizing point" – at some fixed scale μ_0. Of course, in renormalizable quantum field theories (for example in gauge theories) it is enough to fix only a finite (and usually small) set of quantities at the "normalizing point", then the theory is capable to predict *any* cross-sections. However, this circumstance does not abolish this principle feature of quantum field theory, especially transparent in condenced matter physics, where a natural "cutoff" exists (say the scale of elementary atomic lattice) and it is possible to distinguish between the "macroscopic" quantities, which do not depend upon this scale and the "microscopic" ones. Moreover, in the condenced matter physics usually only the *relations* between the microscopic quantities but not the quantities themselves do not depend on the lattice cutoff, and this situation is very similar to what one gets in quantum field theory.

The simplest example is energy of any state, which is defined already in free field theory not as an absolute quantity, but compared to, say, "vacuum energy". Naively the "vacuum energy" gets a contribution from the infinitely many vacuum energies of harmonic oscillators

$$E_{vac} \propto \frac{\hbar}{2}\int d\mathbf{p}\omega(\mathbf{p}) = \frac{\hbar}{2}\int d\mathbf{p}\sqrt{\mathbf{p}^2+m^2} \tag{2.16}$$

In field theory without gravity this quantity is *not* observable and can be considered as a reference point, i.e. one may put, say $E_{vac}=0$. When including gravity, according to the principle of equivalence the vacuum energy is a source for gravitational field. The field theoretical expression (2.16) gives a value absolutely uncomparable to the value of the cosmological constant with any cutoff (or, better, with any scale of supersymmetry breaking). The fundamental theory containing gravity must know how to compute "absolute" values, and this means that such theory in principle cannot be quantum field theory. The problem of vacuum energy or the cosmological constant is one of the principle unsolved problems of modern physics and we will come back to the questions not once below.

From the structure of corrections (2.15) it is clear that at small distances $l^{-1} \sim M_{\mathrm{Pl}} = \gamma_{\mathrm{N}}^{\frac{1}{2-D}}$ gravity, generally speaking, becomes strong. The problems of strong gravitation interaction and related issues of strong gravitational fields, say, in black holes, are even less studied that the problems of strongly coupled gauge theories. One of the well-known effects from the theory of black holes is the linear relation $S=\frac{\mathrm{Area}}{4\gamma_{\mathrm{N}}}$ between the number of states or entropy S and *area* of the horizon (Area) of a black hole [64]. This statement cardinally contradicts the expectations of quantum field theory, where the number of states is always proportional to the volume (but not to the area). This is a kind of indirect argument in favor of the point of view that in strong gravitational fields one may find some fundamental one-dimensional structures; for a detailed discussion of this issue see [29]. Of course, not being decisive, this is one of the indirect arguments in favor of string theory.

3 Main Principles of String Theory

In order to get a consistent theory of quantum gravity one should crucially change the theory at Planckian scales and replace the pointlike objects by one dimensional extended objects – strings. String theory by definition possesses a dimensional constant, which for historical reasons (see formula (3.7) below) is denoted as α'. This constant has dimension of the *square* of length. In "fundamental" string theory, pretending to be the theory of quantum gravity, this parameter can be nothing else but the Planck length, i.e. $\sqrt{\alpha'} \sim 10^{-33}$ cm. However, more generally, its value may be chosen depending on problem under consideration. For example, in string theory applied to the theory of strong interaction at large distances this parameter should be of the order of the hadron size 10^{-13} cm.

Let us point out that α' is the *only* constant, put "by hand" into string theory. It has a clear sense of the *scale* where stringy effects become essential. There are no other constants in string theory, even the dimensionless string coupling g_{str}, as we see below, is not really a parameter, but is rather related to the vacuum condensate of a background field – the so called dilaton. In other words, this constant is a dynamical parameter of the theory.

String theory drastically differs from quantum field theory. We will be coming back to the discussion of this issue many times, so let us now briefly formulate the main points. In string and field theory:

- there is a different "counting in loops", i.e. in field theory and string theory the intermediate state propagating along the loops are counted with different weight factors;
- there is an essential difference in how dimensional reduction looks like, moreover, these theories are especially different in space-times with compact directions;
- space-time shows up in field theory and string theory in totally different ways; string theory is characterized by a "dynamical" nature of space-time. In particular there exist, say, "mirror pairs", i.e. the manifolds which are *not distinct* by string theory;

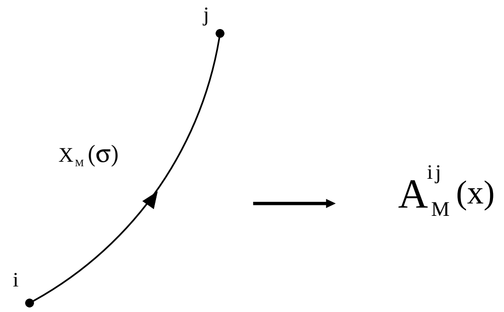

Figure 3: *Massless non-Abelian vector field as a string with quarks at the ends*

- locality and causality also appear in different ways.

As was first noticed by Scherk and Schwarz [50], string theory naturally leads to unification of gauge fields and gravity into one single theory, since in the spectrum of string one automatically gets *massless* vector fields together with massless fields of spin two.

3.1 Gauge Fields and Gravitons

Let us start the discussion of foundations of string theory from an old observation that the theory of one-dimensional extended objects naturally contains vector fields and gravitons. The simplest (though not the most strict) way to see this is to consider a string field or a functional of string contour $\Phi[X_\mu(\sigma)]$ and its expansion in string harmonics (with the Fourier coefficients α_n^μ)

$$X_\mu(\sigma) = x_\mu + \sum_{n\neq 0} \frac{\alpha_{-n}^\mu}{n} \, \exp(in\sigma) \tag{3.1}$$

This expansion obviously has the following form

$$\Phi[X_\mu(\sigma)] = \phi(x) + A_\mu(x)\alpha_{-1}^\mu + \ldots \tag{3.2}$$

After quantization $[\alpha_n^\mu, \alpha_m^\nu] = n\delta_{n+m,0}\delta^{\mu\nu}$ the Fourier coefficients turn into the creation and annihilation operators of string excitations. Then formula (3.2) can be better thought of as the action of the *operator* $\Phi[X_\mu(\sigma)]$ on the Fock vacuum $|0\rangle$ in the space of states of an open string. The first term means that vacuum corresponds to the wave function of a scalar field $\phi(x)$, the next neighbor state $\alpha_{-1}^\mu|0\rangle$ is related to the vector field $A_\mu(x)$. In expansion (3.2) one may take into account only the string harmonics (the coefficients of decomposition in (3.1)) α_n^μ with $n < 0$ (creation operators), since $\alpha_n^\mu|0\rangle = 0$ when $n > 0$.

String quantum mechanics and requirement of invariance under reparameterizations of "internal" co-ordinates on the world-sheet immediately leads to the condition that the vector field $A_\mu(x)$ should be massless. The simplest explanation of this fact is that reparameterizations of co-ordinates on world-sheet have "eaten up" two degrees of freedom, so that physical degrees of freedom are only the *transverse* excitations, say α_{-1}^i, $i = 1, \ldots, D-2$, if speaking about the vector field. Hence, the vector has only $D-2$ physical components, where D is the space-time dimension. This automatically means that the vector field is massless or gauge field, since a massive vector must have $D-1$ physical components. More strictly it can be demonstrated considering the operator of string mass or energy of string excitations

$$M^2 = \frac{1}{\alpha'}\left(\sum_{n=1}^{\infty} \alpha_n^i \alpha_{-n}^i - 1\right) \tag{3.3}$$

which shows that the string spectrum contains the massless gauge field. However, this spectrum starts from the tachyon $\phi(x)$, resulting in additional problems; one of the most effective tools to overcome this problem is supersymmetry.

In order to make vector field $A_\mu(x)$ non-Abelian one should assign the extra indices to the ends of string [74] (for example, of the quark- or antiquark- fundamental representations). Then the vector field becomes matrix $\|\mathbf{A}^{ij}\|$ transforming under *adjoint* representation of the corresponding gauge group (see fig. 3). For quite a long period of time this procedure was performed "by hand" (amplitudes were simply assigned by the Chan-Paton

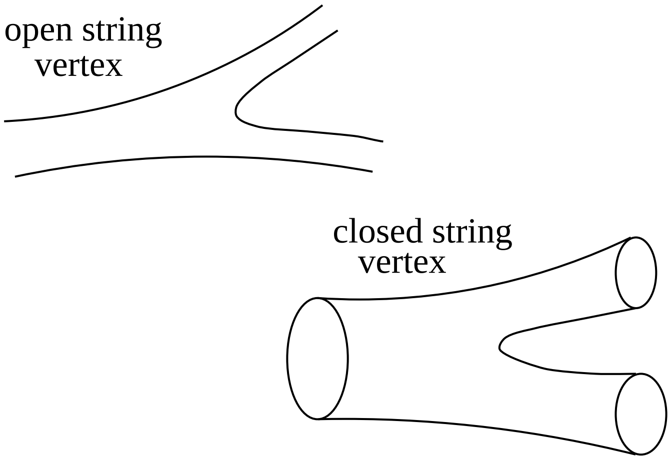

Figure 4: *Interaction vertices of open (at the top) and closed (in the bottom) strings*

factors), until it finally has become clear that a non-Abelian theory naturally arises if one allows existence of so called D-branes (see sect. 4.4). Since it is massless vector field which appears in string spectrum, one gets exactly gauge quantum field theories in the field theory limit $\alpha' \to 0$, when masses (3.3) of all other string harmonics $M^2 \sim \frac{N}{\alpha'}$ (with N being the eigenvalue of the operator $\sum \alpha^i_n \alpha^i_{-n}$ of the "number of particles" – string harmonics in formula (3.3)) become very large and their excitations in low-energy effective theory, i.e. at the distances much larger than $\sqrt{\alpha'}$ can be neglected.

In supersymmetric string theory the massless sector contains vector supermultiplets, where the rest of the states are constructed by supersymmetry. In the low-energy limit this leads to a supersymmetric theory of the Yang-Mills fields as an effective theory of massless modes over the possible vacua of string theory. According to modern general philosophy quantum field theories (in particular, supersymmetric gauge theories got in this way) can be considered as an effective description of physics near different vacua of string theory. These vacua can be related to each other by duality transformations – some *discrete* transformations, exchanging different vacua of string theory and, therefore, different quantum field theories.

The expansion over modes of a closed string is similar to formula (3.2) but since the interaction (say with the background fields) in the closed sector takes place over the whole world-sheet, one should consider two sets of string harmonics corresponding to left and right waves independently propagating over the string world sheet. These waves are solutions to the equations of motion of free string: $\alpha^\mu_n e^{in(\tau+\sigma)}$ Ë $\tilde{\alpha}^\mu_n e^{in(\tau-\sigma)}$. The spectrum again starts from the tachyon (the different one with the modulus of mass squared twice that of the tachyon of the open string spectrum). Massless fields correspond to the states $\alpha^\mu_{-1}\tilde{\alpha}^\nu_{-1}|0\rangle$, or, more exactly to their linear combination. Dividing the second rank tensor into irreducible representations of the Lorentz group, it is easy to see that the corresponding fields consist of, first

$$\left(\alpha^\mu_{-1}\tilde{\alpha}^\nu_{-1} - (\mu \leftrightarrow \nu)\right)|0\rangle \cdot B_{\mu\nu}(x) \tag{3.4}$$

or the antisymmetric tensor field $B_{\mu\nu}$, second

$$\delta_{\mu\nu}\alpha^\mu_{-1}\tilde{\alpha}^\nu_{-1}|0\rangle \cdot \varphi(x) \tag{3.5}$$

i.e. the massless scalar usually called a dilaton. It is the vacuum value of the dilaton which gives the value of string coupling constant. Finally, the rest of the components form the massless and traceless symmetric tensor

$$\left(\alpha^\mu_{-1}\tilde{\alpha}^\nu_{-1} - \frac{1}{D}\delta_{\mu\nu}\alpha^\lambda_{-1}\tilde{\alpha}^\lambda_{-1}\right)|0\rangle \cdot G_{\mu\nu}(x) \tag{3.6}$$

or the *graviton*.

All considerations of this section are based by now on the simplest quantum mechanics of the free string. Switching on the interaction (see fig. 4), one may easily verify the two following important properties of the theory:

- Tree amplitudes of scattering of massless states of open strings in the limit $\alpha' \to 0$ turn into the scattering amplitudes of vector gauge bosons, and similar the scattering amplitudes of the states (3.6) of the closed sector turn into the amplitudes of graviton scattering [50].

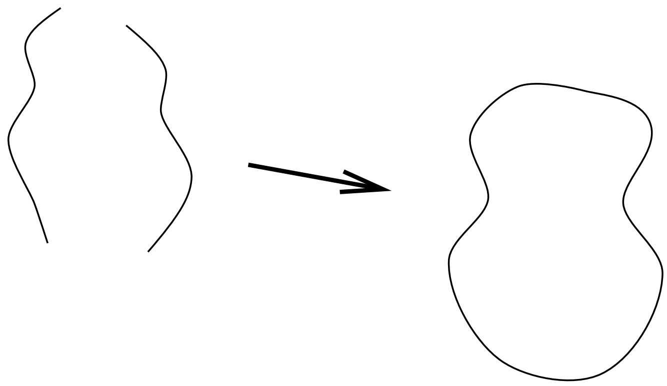

Figure 5: *Interaction of open strings leads to closed strings*

- Interaction of two open strings leads to appearance of the closed strings (see fig. 5). Together with the previous remark this means that gauge field theories, constructed in the framework of string theory, necessarily lead to the appearance of gravity.

3.2 Massive Fields and Ultraviolet Cutoff

Let us turn now to massive fields of string spectrum, Their masses M (see (3.3)) are measured in units of the (inverse) string length or the Planck mass $\sqrt{n/\alpha'}$, where n is the number of corresponding string harmonics or excitation level. It is easy to understand that this number is linearly related to the (maximal possible) spin of the excitation J. The exact relation can be written in the form of so called Regge trajectory – the linear function [9]

$$J = \alpha(M^2) \equiv \alpha_0 + \alpha' M^2 \tag{3.7}$$

and from (3.3) it immediately follows that $\alpha_0 = 1$ for an open string. The relation between spin and mass (3.7) was known long ago in the theory of strong interactions, which after the works of Veneziano [46], Nambu and Goto [47] became a "parent" of string theory. Notice immediately that all excitations with higher spins in string theory do have masses of the order of the Planck mass. Therefore their absence in visible spectrum does not contradict to their presence in the theory, unlike of the non-removable well-known defect of the quantum field theories with higher spins.

In the limit $\alpha' \to 0$ string theory reproduces the theory of pointlike particles. From the whole "tower" of fields only the massless fields survive (under assumption that the tachyon problem is solved; this problem will be discussed in detail in sect. 3.5 and 6.3). The size of a string can be estimated, for example, computing the correlator

$$\begin{gathered}\langle 0| \int_{d\sigma} (X(\sigma) - x)^2 |0\rangle = \alpha' \sum_{n>0} \frac{1}{n^2} \langle 0|\alpha_n \alpha_{-n}|0\rangle \propto \alpha' \sum_{n>0} \frac{1}{n} \propto \\ \propto \alpha' \log n_{max} = \alpha' \log(\sqrt{\alpha'} E_{max})\end{gathered} \tag{3.8}$$

where n_{max} and E_{max} are "number" and energy of maximal excited string harmonic. This formula shows that the size of string is of the order of $\sqrt{\alpha'}$ (it grows very slowly with energy), which justifies the interpretation of the only dimensional parameter of string theory α' as a square of string length.

Notice finally, that the number of quantum states in the string spectrum grows rapidly with the energy of excitations. At large energies the spectral density behaves as [10]

$$\rho(M) \propto \exp(2\pi\sqrt{\alpha'} M) \tag{3.9}$$

This behavior leads to absolutely unusual (and different from quantum field theory) properties of string theory at small distances or large energies – i.e. at the Planck scale.

[9] Let us stress once again that dimensional parameter α' characterizes the scale when string effects become to be essential. Therefore the exact value of this quantity is different for strings, arising as effective description of strong interactions at large distances and "fundamental" strings, corresponding to quantum gravity. Using the notation originally introduced in the context of hadron physics, we will consider however, if the opposite is not stated directly, this parameter to be equal to square of the Planck length.

[10] The numeric coefficient in front of $\sqrt{\alpha'} M$ in the formula (3.9), generally speaking, depends on the particular string model. Literally in (3.9) it is written as in the theory of closed strings, where it is maximally universal. One of the simplest methods to derive this coefficient for any string model is to consider the singularities of string propagators [95].

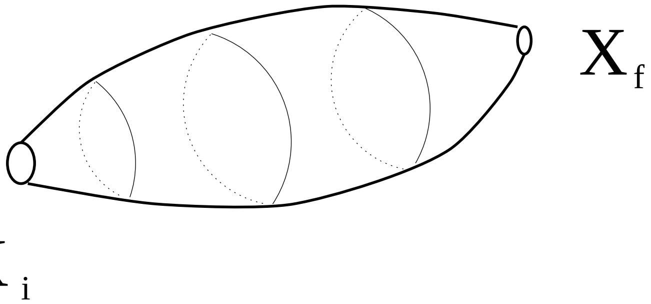

Figure 6: *Propagator of closed string with fixed boundary contours. Choosing these contours as points the propagator becomes a function of two variables $G(X_f, X_i)$, and can be compared to a similar object in quantum field theory.*

- One of the ways to see this already in the theory of non-interacting strings is to consider the thermodynamics of string states. Neglecting interaction free energy has the form

$$F(\beta) \sim \int dE\rho(E)\exp(-\beta E) \tag{3.10}$$

  and for the density of string states (3.9) this integral converges only at $\beta > \beta_H = 2\pi\sqrt{\alpha'}$ or at the temperatures less than the Hagedorn temperature $T_H = \frac{1}{\beta_H} = \frac{1}{2\pi\sqrt{\alpha'}}$ [11]. It means that at the Hagedorn temperature the phase transition is possible [89]. Simple calculations show that at hight temperatures the number of (gauge-invariant) states in string theory is much less than in quantum field theory. For the "normalized" free energy in string theory independently of space-time dimension D one has $\frac{F}{VT} \underset{T\to\infty}{\propto} T$ instead of $\frac{F}{VT} \underset{T\to\infty}{\propto} T^{D-1}$ in field theory. Not being yet finally understood, this property demonstrates the qualitative agreement between the high-energy properties of string theories with corresponding (hypothetical) properties of gravity.

- Another manifestation of the same effect is violation of "microlocality" in string theory, related to the growth of spectral density according to (3.9). Computing the Green function or propagator of string between "pointlike" initial and final states (see fig. 6), and studying its singularities it is easy to see that they look like singularities of *non localizable* theory, i.e. lie within some hyperboloid getting into the distance of the order $\sqrt{\alpha'}$ into the space-like region [90] (see fig. 7).

- Scattering amplitudes in string theory at large energies crucially differ from the corresponding amplitudes in quantum field theory by softer behaivior, this can be seen already at the level of the Veneziano amplitude (see, for example, [2]). Due to summing over infinitely many states in the intermediate channels, the amplitudes of string theory contain the "cutting" factor at high energies.

Notice finally, that the opposite limit to field theory $\alpha' \to \infty$ (the so called "nill-strings") is very singular. Being a complicated technical problem, this limit is most likely senseless from the physical point of view. It corresponds to the theory at the energies much more than Planckian, i.e. in the region where neither field theory nor even string theory are literally applicable and taking such limit is similar to an attempt to use field theory beyond the scale of ultraviolet cutoff.

3.3 String Perturbation Theory – Sum over Two-dimensional Geometries

The perturbative structure of string theory can be defined by the "loop expansion", see fig. 8,

$$\mathcal{F} = \sum_{g=0}^{\infty} g_{\rm str}^{2g-2} F_g \tag{3.11}$$

or by expansion over topologies or *genera* of the world sheets being two-dimensional Riemann surfaces. The role of parameter of this expansion is played by$g_{\rm str}$ – the string coupling constant. Notice immediately that expansion (3.11) is written for the free energy or the *logarithm* of the partition function (in contrast to quantum

[11] The Hagedorn temperature coincides with the Hawking temperature of the black hole whose gravitational radius is equal to string length $M\gamma_{\rm N} \sim \sqrt{\alpha'}$.

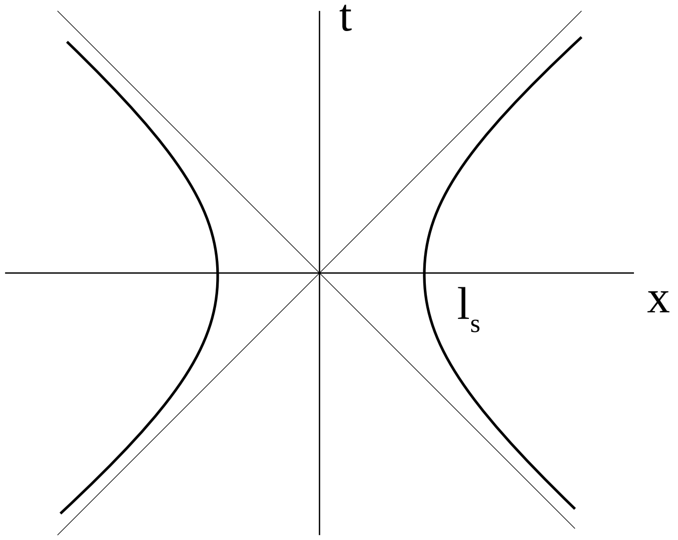

Figure 7: *The singularities of propagator in string theory. In contrast to local quantum field theory they are located inside the hyperboloid, getting into the distances of $l_s = 2\pi\sqrt{\alpha'}$ into the space-like domain.*

field theory) since it includes summation only over "connected diagrams". Literally the loop expansion on fig. 8 is valid for the theories with only closed strings. These theories include the closed bosonic strings as well as so called superstrings of type II [12], on which we will mostly concentrate below. If the theory contains open strings together with closed one should also add the world-sheets with boundaries.

Let us also note that the normalization in (3.11) as well as in fig. 8 is chosen in such way that the contribution of any genus is proportional to the particular power of string coupling $g_{\rm str}$, which is equal, up to a sign, to the *Euler characteristic* of the corresponding world-sheet. Due to this normalization the expansion starts with $g_{\rm str}^{-2}$ and includes for closed strings only *even* powers of string coupling. In the theory of open strings for the world-sheets with boundaries one would also get the odd degrees of the coupling constant. This means that the string coupling in closed sector is in fact proportional to the square of the open string coupling and this fact will be important below when discussing the nonperturbative theory.

The contribution of each genus is computed by the Polyakov path integral [52] over the string co-ordinates and two-dimensional geometries or metrics on world-sheet [13]

$$F_g = \int Dh_{ab} D\mathbf{X} \exp\left(-\int_{\Sigma_g} \partial\mathbf{X}\bar{\partial}\mathbf{X}\right) \tag{3.12}$$

where $\mathbf{X}$ are co-ordinates of string, being at the same time from the point of view of two-dimensional world sheet theory the fields of a free field theory, and h_{ab} denote metrics on Riemann surface Σ_g of genus g. The summation over two-dimensional geometries in (3.12) was originally formulated by Polyakov as integration over metrics. If, however, one takes into account the invariance under reparameterizations on world sheets, the sum is really taken over the "equivalence classes" of metrics (with respect to changes of co-ordinates or reparameterizations) and it is these equivalence classes which correspond to physically different configurations.

On the first glance the action $\int_{\Sigma_g} \partial\mathbf{X}\bar{\partial}\mathbf{X}$ in formula (3.12) does not at all depend on two-dimensional metric h_{ab}. Two out of three its components may be immediately "killed" by two reparameterizations of the world-sheet co-ordinates, say, the metric can be brought by reparameterizations to the conformal form $h_{ab} = \exp(\varphi)\delta_{ab}$, when it is determined by a single function φ on string world-sheet. It is easy to check that free action (3.12) does not at all depend upon the "conformal factor" φ and the integral over metrics in (3.12) looks like being trivial. However, this is not true. The reason is that two-dimensional theory (3.12) is a simple quantum mechanics but with *infinitely* many degrees of freedom and therefore the integral in (3.12) should be regularized.

If we require that regularized theory should be independent of the choice of co-ordinates on string world-sheet (and such arequirement is absolutely necessary from physical point of view – the sensible physical theory must

[12] When D-branes are absent, see sect. 4.3 and 4.4.

[13] Since string theory *by definition* contains an integral over two-dimensional metrics it is often identified with two-dimensional quantum gravity. Indeed the parallels between string theory and quantum gravity in two dimensions are very useful for studying both theories. However, one should remember the principle difference in space-time interpretation, which for string theory is multidimensional and the observables in string theory are defined in multidimensional space-time.

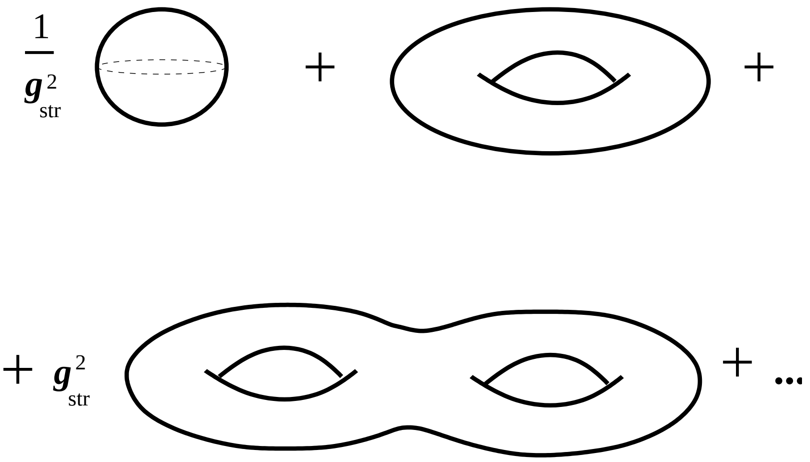

Figure 8: *String "Feynman diagrams" corresponding to the first three terms of the perturbative expansion (3.11) for closed strings. The tree-level contribution (of the order of* $\frac{1}{g^2_{\mathrm{str}}}$ *in "string normalization") (3.11)) corresponds to the sphere, the one loop contribution is given by the torus, the two-loop by the figure of eight, etc.*

not depend on co-ordinates on unobservable world-sheet of the Planckian size) the regularization (for example, cutoff) should be introduced *covariantly*. This means that *quantum* theory (3.12) in general does depend on metric h_{ab} or at least on its conformal class. Such phenomenon is called as anomaly (see, for example, the review [34] and references therein), and in our case we deal with two-dimensional conformal or gravitational anomaly. Calculation of this anomaly in [52] has demonstrated that two-dimensional geometry essentially restricts the properties of space-time which is a target-space for string theory. The origin of these restrictions is that contribution to anomaly of "physical degrees of freedom" should be compensated by the contribution of two-dimensional geometry (supergeometry) itself. And it is this constraint which leads to well-known critical dimensions $D = 26$ (or $D = 10$) "fixed by God". Such restrictions are not as strong as one had thought at the beginning of string era, but nevertheless string theory in some sense chooses the space-time "itself". The space-time in string theory should be essentially multidimensional, though partially these dimensions can be "small" – i.e. responsible for the "internal" degrees of freedom in spirit of the Kaluza-Klein models [62].

The computation of anomaly [52] shows that in quantum theory the conformal factor φ "alives" and acquires the meaning of extra (singled out) co-ordinate of the space-time. Anomaly adds the kinetic term for the field φ to the action (3.12), so that (in flat space-time) the total action acquires the form

$$\int_{\Sigma} \left(\partial \mathbf{X} \bar{\partial} \mathbf{X} + \partial \varphi \bar{\partial} \varphi + \ldots \right) \tag{3.13}$$

where in some natural normalization the field φ should be regarded as imaginary. In other words formula (3.13) is naturally interpreted as a free action in Minkowski space. The interpretation of time as "scale factor" arising in the framework of string theory is a bit similar to analogous interpretation in general context of gravity and cosmology.

Let us return to the properties of the path integral (3.12) over two-dimensional geometries. In the case of pointlike particles this integral is reduced to the finite-dimensional integral over the Feynman parameters, which have the meaning of invariant lengths of the trajectories of particles. In such away the Feynman diagrams (say, in the ϕ^3-theory) arise directly at the first-quantized level. The main physical problem coming out of the integrals over Feynman parameters (and hence from the integral over one-dimensional geometries) is the appearance of ultraviolet divergencies due to contributions of trajectories of infinitely small lengths. In string theory these singularities are naturally regularized when one passes from world-lines intersecting at some points to smooth world sheets (this immediately leads to the fact that only cubic interaction is possible in string theory).

A more delicate effect is that two-dimensional geometry *regularizes* the contribution of small distances since this contribution is geometrically equivalent to the contribution of trajectories of large lengths. According to the main principle of quantum physics the summation should be taken only over the independent configurations. One should immediately conclude that in order to avoid "double counting" the contribution of the trajectories with small lengths should *not* be counted at all, if all equivalent "infrared" configurations are already taken into account. As a result of this logic we get a striking consequence that in string theory Γб«Ё гзвΓл ўбΓ нЄўЁў«ΓвлΓ Ё ЄдЁJгажЁЁ ў *by definition* the ultraviolet problems of the quantum field theory are absent, more strictly there are no ultraviolet divergencies if there are no infrared [14]. This statement follows from the analysis of finite-dimensional part of the integral over two-dimensional geometries given by the integral over moduli spaces of complex structures of Riemann surfaces (this issue is in the center of discussion of the main part of review [9]).

According to the Belavin-Knizhnik theorem [58] the integral over metrics (3.12) is reduced to the integral over the moduli space of complex structures of the Riemann surfaces

$$F_g = \int_{\mathcal{M}_g} d\mu(y)|f(y)|^2 \tag{3.14}$$

where $\mathcal{M}_g$ is the (finite-dimensional) moduli space of complex structures of the Riemann surface Σ_g. The concrete choice of the integration measure depends on particular choice of a string model, for the bosonic string this is the Mumford measure [58]. It is the modular invariance of the integrand in (3.14) leading to the fact that contributions of the trajectories of small lengths and the trajectories of large lengths are physically equivalent. The formulation (3.12), (3.14) allows one in principle to use the symmetry properties in order to get some nonperturbative information, though by its own definition this is just a perturbative expansion around some vacuum and the integral (3.12) computes only the g-loop correction of the expansion of string perturbation theory.

3.4 Dynamical Nature of Space-Time and Two-dimensional Conformal Theories

Let us come back to the fact that the contribution of the new co-ordinate coming from two-dimensional metric allows to cancel the conformal anomaly. This condition is not empty (in the sense that it does not take place everywhere) and leads to *dynamical restrictions* on the properties of physical space-time. The basic restrictions look as follows:

- In flat space-times string theory exists only in some *distinguished* or critical dimensions. The simplest bosonic string (3.12), (3.13) demands the total number of dimensions to be $D = 26$ (including time), and the theory of fermionic or supersymmetric strings (two-dimensional supergravity) fixes the critical dimension to be $D = 10$.

- In nontrivial background fields, say, when metric is not flat, the background fields should satisfy the classical equations of motion, in particular the Einstein equations

$$R_{MN}(G) - \frac{1}{2}G_{MN}R(G) - T_{MN} = \mathcal{O}(\alpha') \tag{3.15}$$

  up to the string corrections. In eq. (3.15) $G_{MN} = G_{MN}(X)$ is the space-time metric, $R_{MN}(G)$ is its Ricci tensor and T_{MN} is the stress-energy tensor of the other background fields. Moreover, in presence of nontrivial background fields the anomaly cancellation condition is changed. In such case the critical dimension ($D = 26$ or $D = 10$) "moves", i.e. changes due to contribution of corrections in α' to the anomaly – the terms, starting with $\alpha' R(G)$.

Generally speaking, the space-time should not be necessarily Minkowski space or the Euclidean flat space [15], say $\mathbf{R}^4$, it may have a nontrivial metric (satisfying the Einstein equations due to the two-dimensional symmetries [57]). It can be even a nontrivial compact manifold (or, more exactly has a compact part), corresponding, as already mentioned above, to the internal (gauge) degrees of freedom in spirit of the Kaluza-Klein models. The Polyakov path integral (3.12) should be then understood in "generalized" sense when instead of free infinite-dimensional quantum mechanics (or two-dimensional field theory (3.12) with the fields X, to be interpreted as space-time co-ordinates) one should deal with some generic sigma-model

$$\int_\Sigma \left(G_{MN}(X)\partial X^M \bar\partial X^N + \mathcal{R}^{(2)}\Phi(X) + \ldots \right) \tag{3.16}$$

where $\mathcal{R}^{(2)} = \mathcal{R}^{(2)}(h)$ is the curvature of the two-dimensional metric, while $G_{MN}(X)$ and $\Phi(X)$ are nontrivial background fields for the space-time metric and dilaton. A principal new moment in string theory is that the theory "adjusts" to itself the space-time where it exists. More strictly, it imposes essential constraints on the characteristics of the target space-time and forces the background fields to be solutions to the equations of motion. Let us also point out that comparing eqs. (3.16) and (3.11) and using the Gauss-Bonnet theorem $\int_\Sigma \mathcal{R}^{(2)}(h) = 2-2g$, (where $g = g(\Sigma)$ is genus of the Riemann surface Σ) one gets the relation between the "zero

[14] This is not the case for many string models due to presence of tachyons.

[15] The problems of signature of space-time are still beyond the framework of string theory and we will not discuss it here. Let us only point out that we imply everywhere a possibility of smooth analytic continuation of the theory in Minkowski space to the Euclidean space and we will not distinguish between these two formulations below.

mode" Φ_0 of the dilaton field $\Phi(X)$ (more exactly of its vacuum expectation value) and the string coupling constant $g_{\rm str} = \langle\exp(\Phi_0)\rangle$.

Considering string theory in the external background fields, including nontrivial metric of the space-time (such theories for historical reasons are usually called as two-dimensional sigma-models), it is necessary all the time to look after the condition of conformal invariance, which is reminiscent of the reparameterization invariance after the metric h_{ab} has been chosen in conformal form, see (3.12), (3.13) and (3.16)). In other words, nontrivial background fields should necessarily correspond to the two-dimensional conformal sigma-models, or, more directly to the two-dimensional conformal theories [53]. The difference between these two notions is only in the fact that majority of known two-dimensional conformal theories have only an approximate description in sigma-model terms. Usually, an explicitly known nontrivial sigma-model can correspond only to "bare" values of the background fields, while the exact background fields, which hypothetically describe the exact conformal theory are not really known. In such a case the conformal field theory can nevertheless be formulated axiomatically [53] or, in terms of free field theories [80, 81, 96], corresponding to the simplest dilaton background [16].

Two-dimensional conformal field theories [53] are the theories with invariance under the action of the *infinite-dimensional* (only in two-dimensions!) group of conformal symmetry. This group if formed by holomorphic reparameterizations on world sheets, keeping metric in conformal form $h_{ab} = \exp(\varphi)\delta_{ab}$. The generators of such transformations form the Virasoro algebra

$$[\mathcal{L}_n, \mathcal{L}_m] = (n-m)\mathcal{L}_{n+m} + \frac{c}{12}\delta_{n+m,0} \tag{3.17}$$

and in the "classical" case (at $c=0$) may be represented as $\mathcal{L}_n = -z^{n+1}\frac{d}{dz}$, i.e. form the basis of holomorphic vector fields on the world-sheet Σ, parameterized by complex co-ordinates $(z,\bar z)$. Implying that conformal symmetry is an exact symmetry of quantum theory (and this is again a natural requirement of independence of physics of the choice of co-ordinates on the world-sheet of the Planckian size), one gets immediately an infinite number of constraints (the Ward identities) on the correlation functions in two-dimensional theory [53]. This allows in principle to calculate any two-dimensional correlator, being the "building blocks" for string amplitudes. It turns out that the same statement can be formulated alternatively: despite all conformal theories corresponding to nontrivial manifolds in space-time being not free theories (3.13) in the literal sense, for any conformal theory there exists a representation in terms of free fields or so called *bosonization* [80, 81, 96]. This means that in any nontrivial space-time, consistent with two-dimensional conformal invariance, string theory is *in principal defined* perturbatively and the integrals (3.12) and (3.14) can be calculated. Bosonization effectively reduces the computations in nontrivial conformal theories to the calculation (of quite complicated correlation functions) in the theories with quadratic action

$$S_{CFT}(\varphi) = \int_\Sigma \left(\partial\varphi\bar\partial\varphi + \alpha_0\mathcal{R}^{(2)}\varphi\right) \tag{3.18}$$

where the constant α_0 (or constant vector in case of many fields) is related to the central charge $c_{CFT} = 1 - 12{\alpha_0}^2$. This is the way how non-integer central charges of nontrivial theories arise from the free theories with central charges just equal to the number of fields, $c = D$. It is also useful, as follows from comparison with (3.16), to interpret action (3.18) as the action of a string in the external *linear* dilaton background $\Phi(\varphi) = \alpha_0\varphi$. We will see below that such a background is also singled out in string theory also from other points of view.

Besides, for generic conformal theories one should specially notice that a single conformal theory may correspond in general to strings on *different* manifolds $\mathcal{X}_1$ and $\mathcal{X}_2$. Such manifolds are called mirror manifolds [23, 24]. The simplest example is a free theory of a field, taking values on a circle – the theories on circles $\mathcal{X}_1 = S_R$ of radius R and $\mathcal{X}_2 = S_{\alpha'/R}$ with the radius α'/R are equivalent, see sect. 4.2).

Let us recall once more that the amplitudes in string theory are built from the correlation functions of two-dimensional conformal field theory. More exactly, the scattering amplitudes of, say, massless excitations above some vacuum do correspond to the particular correlators in two-dimensional conformal field theory corresponding to this vacuum. These operators are fixed by the set of corresponding quantum numbers and by condition of conformal invariance – the consequence of reparameterization invariance on the world-sheet. It is remarkable that conformal invariance immediately leads to all physical requirements on the operators of physical particles. Let us demonstrate this on the example of the operator of emission or absorbtion of a photon (in a flat space-time)

$$\epsilon\cdot\partial X\exp(ip\cdot X) \tag{3.19}$$

with momentum p and polarization vector ϵ. First, conformal invariance says that a "physical operator" must have unit dimension, then and only then the result of integration over the boundary of the world-sheet (in case of open strings, or over the whole world-sheet in case of closed strings) will not depend on the choice of co-ordinates. For the operator of photon (3.19) it means (due to unit dimension of pre-exponent) that $p^2 = 0$, or, alternatively, that the (*anomalous* in the sense of two-dimensional conformal field theory) dimension of the exponent in (3.19) vanishes. Thus, from the condition of *two-dimensional conformal invariance* one immediately obtains that the photon is massless. In fact this derivation is just a little bit more strict variant of the argumentation from the beginning of sect. 3.1.

[16] A nice exception consists of two-dimensional sigma-models on group manifolds and conformal theories corresponding to them [54]. However, even in this case it is simpler and more natural to construct the conformal theory just *requiring* that conformal symmetry is an exact quantum symmetry consistent with the current algebra, always existing on group manifold [55].

Slightly more detailed analysis of the conformal invariance leads rapidly to the transversality of physical photon $\epsilon \cdot p = 0$, or to *gauge* invariance. Indeed, decomposing the polarization vector into the transversal and longitude parts $\epsilon_M = \epsilon^{\perp}_M + \epsilon^{\|}_M$, so that $\epsilon^{\perp} \cdot p = 0$ Ë $\epsilon^{\|}_M \propto p_M$, one easily finds that

$$\epsilon^{\|} \cdot \partial X \exp(ip \cdot X) \propto p \cdot \partial X \exp(ip \cdot X) \propto \partial\left(\exp(ip \cdot X)\right) = \mathcal{L}_{-1} \cdot \exp(ip \cdot X) \tag{3.20}$$

i.e. the contribution of the longitudinal part is the total derivative and disappears after the integration over the boundary of the world sheet. In other words, using the last equality in (3.20), one may say that the operators or states corresponding to physical particles are defined in the language of two-dimensional conformal theories up to the "gauge" states of the form $\mathcal{L}_{-1}|\Psi\rangle$ and with the vanishing norm. Thus, the "ghost-free" requirement of two-dimensional theory leads to the gauge invariance in physical string spectrum.

3.5 Supersymmetry and Fermions

Let us now briefly discuss the extra world-sheet fields and related internal degrees of freedom. One of the important properties of string theory is that by introducing supersymmetry on world-sheet one immediately obtains the space-time fermions [17].

Already in the degenerate example of string – the relativistic particle – it is enough to introduce the world-line supersymmetry [77], to get the space-time fermions. The world-line action can be defined requiring the invariance under the (one-dimensional!) supersymmetry with Grassmann parameter ϵ

$$\begin{aligned} \delta X &= \epsilon\Psi \\ \delta\Psi &= -\epsilon\left(\dot{X} + \frac{1}{2}\chi\Psi\right)e^{-1} \\ \delta\chi &= -2\dot{\epsilon} \\ \delta e &= -\epsilon\chi \end{aligned} \tag{3.21}$$

The corresponding invariant action

$$\frac{1}{2}\int_{dt}\left(\frac{\dot{X}^2}{e} + \Psi\dot{\Psi} + \frac{\chi}{e}\Psi\dot{X} + m^2(e + \frac{1}{4}\chi d_t^{-1}\chi)\right) \tag{3.22}$$

includes in addition to co-ordinates X_M and one-dimensional "metric" e the Grassmann "gravitino" χ and fermionic variables Ψ_M with the first-order kinetic term, such that these variables coincide with their own momenta $\Psi_M = \frac{\delta S}{\delta\Psi_M}$. After quantization one gets the relations $[\Psi_M, \Psi_N]_+ = \delta_{MN}$, i.e. the Grassmann variables Ψ_M turn into the Dirac gamma-matrices and the wave function carries now also the *space-time* spinor index, since it becomes a vector of certain representation of the Clifford algebra. The corresponding representation in terms of the (one-dimensional analog) of the Polyakov path integral with the action (3.22) allows to compute Green functions in the theory of Dirac fermion.

Notice, that the world-line supersymmetry (3.21) (as well as its direct generalization – the supersymmetry on the string world-sheet) is practically identical to the well-known supersymmetry in quantum mechanics. The simplest example of supersymmetry in quantum mechanics is a particle in magnetic field, which can be considered as a quantum mechanical system with the Hamiltonian $H = (\boldsymbol{\sigma} \cdot \mathbf{P})^2$ (with the Pauli matrices $\boldsymbol{\sigma}$ being the simplest representatives of the Dirac matrices). The role of supergenerator is played by the Dirac operator $\boldsymbol{\sigma} \cdot \mathbf{P}$, and this exactly corresponds to the interpretation of supersymmetry transformations as "square roots" of the energy-momentum operators. The essential feature of supersymmetry in quantum mechanics (in particular that of (3.21)) is that the related "fermionic number" is not really "fermionic" from the point of view of space-time.

Indeed, when the role of Hamiltonian is played by the square of the Dirac operator, the "fermionic number" is nothing but direction of spin. Therefore from the perspective of physical space-time the supersymmetric "bosons" and "fermions" just correspond to different directions of spin of a "real space-time" fermion, whose wave function satisfies the Dirac equation. As we see below the world-sheet supersymmetry in string theory reminds one a lot the supersymmetry in quantum mechanics apart from details with the boundary conditions

[17] Here one should make a few extra comments. This property in fact can be detected already at the level of pointlike particles. Moreover, in some sense (without using the notion of supersymmetry) it was known long before the string theory appeared. Nevertheless, it seems to be extremely important that only in string theory or on two-dimensional world-sheets, this property arises naturally and without "pathologies" of the one-dimensional case.

due to a extra co-ordinate on the world-sheet. It is quite nontrivial that this "auxiliary" supersymmetry of a quantum-mechanical type leads to the "real" space-time supersymmetry in string spectrum.

Hence, things are much more interesting for the fermionic string – the first-quantized theory with the world-sheet action

$$\frac{1}{2\pi\alpha'}\int_{\Sigma}\left(\partial X\bar{\partial}X+\Psi\bar{\partial}\Psi+\bar{\Psi}\partial\bar{\Psi}+\chi\Psi\bar{\partial}X+\bar{\chi}\bar{\Psi}\partial X+\frac{1}{2}\bar{\chi}\chi\bar{\Psi}\Psi\right) \tag{3.23}$$

invariant under the transformations of two-dimensional supergravity [78]. First three terms in the expression (3.23) (at $\chi=\bar{\chi}=0$) correspond to the action, invariant under the *global* two-dimesional supersymmetry transformations on world-sheet [75]. Depending on the boundary conditions (periodicity or antiperiodicity or their analogs in the open string case) the fermionic fields Ψ either do not or do contain the "zero mode" – the constant component $\Psi_M^{(0)}$, which in complete analogy with the example of fermionic particle may turn into the set of Dirac matrices after quantization $[\Psi_M^{(0)},\Psi_N^{(0)}]_+=\delta_{MN}$.

Thus, depending on the choice of boundary conditions, there are two sectors in fermionic string. The wave functions of one sector possess an index of a representation of the Clifford algebra and correspond to the space-time fermions, while the wave functions of another sector do not have such indices and correspond to the space-time bosons. The corresponding two-dimensional conformal theory [84, 87] allows to compute the correlation functions, corresponding to arbitrary scattering amplitudes in the fermionic string.

After all that it is natural to ask how the states of the fermionic string spectrum corresponding to space-time bosons and space-time fermions are related to each other. At first glance these two sectors – bosonic and fermionic – differ too much from each other, for example, the bosonic sector (or the Neveu-Schwarz sector [48]) contains tachyon, while the fermionic sector (or the Ramond sector [49]) is tachyon free. Nevertheless, there exists a natural "GSO-projection" (i.e. procedure leaving only half of the states in the spectrum) [51], which results in leaving in the spectrum the equal number of states from both sectors in such a way that the full spectrum (after projection) becomes space-time supersymmetric!

Moreover, at the level of the one-loop partition function this projection arises naturally after summing over all possible boundary conditions of fermionic fields [85]. All this means that supersymmetry on the world-sheets of fermionic strings leads to the supersymmetry in (ten-dimensional) space-time. The resulting theory – the "reduced" fermionic string with ten-dimensional supersymmetry, after John Schwarz is often called superstring.

In the open string sector the GSO-projection leaves in the Neveu-Schwarz sector the subsector with *odd* "fermionic number" (in the sense of world-sheet fermions), for example the massless vector $\Psi^{\mu}_{-1}|0\rangle_{NS}$ is left in the spectrum of open superstring while the naive "vacuum" or the Neveu-Schwarz tachyon $|0\rangle_{NS}$ is "killed" by the GSO-projection. In the Ramond sector the GSO projection leaves only the space-time fermions with fixed chirality (the eigenvalue of the operator $\frac{1}{2}(1\pm "\Gamma_5")$, $"\Gamma_5"\propto\prod_{M=1}^{10}\Gamma_M$, acting on the ten-dimensional Majorana spinors), the number of such fermionic states (at each mass level) is exactly equal to the number of states in the Neveu-Schwarz sector with the odd "fermionic number". Hence, in the theory of closed strings one may have two different superstring theories. One would contain the fermions of different chiralities while the other – the fermions of the same chirality: the first is called a type IIA theory while the second – a theory of type IIB.

It turns out that superstrings can be reformulated without two-dimesional world-sheet Neveu-Schwarz-Ramond type fermions. There exists anaternative Green-Schwarz formulation [79], using the extra Grassmann fields $\theta_\alpha(\sigma,\tau)$ (spinors in ten-dimensional space-time in contrast to the ten-dimensional vectors $\Psi_\mu(\sigma,\tau)$) explicitly invariant under the ten-dimensional supersymmetry transformations. However, the variables $\theta_\alpha(\sigma,\tau)$ behave as scalars with respect to two-dimensional reparameterizations of co-ordinates and two-dimensional supersymmetry is not a symmetry of the Green-Schwarz superstrings.

The investigation of anomalies, started in [56], has brought us to the following list of anomaly-free superstring models: type IIA and type IIB theories (closed string non-chiral and chiral theories with $\mathcal{N}=2$ in ten dimensions), type I theory (which includes open strings) and theories of heterotic strings [82] (the string models where, say, left or holomorphic part corresponds to the twenty-six-dimensional bosonic string with extra compactification while the right or antiholomorphic part – to the ten-dimensional superstring) with the gauge groups $SO(32)$ and $E_8\times E_8$.

Unfortunately the ten-dimensional superstring pretending to be the most successful among existing string models is strictly defined, in general, only at tree and one-loop levels. Starting from the two-loop corrections (the last diagram depicted at fig. 8) to the scattering amplitudes all expressions in the perturbative superstring theory are really *not* defined. The reason for that comes from the well-known problems with supergeometry or integration over the "superpartners" of the moduli of complex structures.

In contrast to the bosonic case (3.14), where the integration measure is fixed by the Belavin-Knizhnik theorem, the definition of the integration measure over supermoduli (or, more strictly, the odd moduli of super-complex structures) is still an unsolved problem [88, 22]. The moduli spaces of the complex structures of Riemann surfaces are non compact, and the integration over such spaces requires special care and additional definitions. In the bosonic case, when the integrals over moduli spaces diverge, the result of integration in (3.14) is defined only up to certain "boundary terms" – the contributions of degenerate Riemann surfaces or the surfaces of lower genera (with less "handles", see fig. 8). In the superstring case one runs into more serious problems since the very notion of the "boundary of moduli space" is *not defined.* Indeed the integral over the Grassmann odd variables does not "know" what is the boundary term. This is the fundamental reason why the integration measure in fermionic string is not well-defined and depends on the "gauge choice" or the particular choice for the "zero modes" χ in the action (3.23). For two-loop contributions this problem can be solved "empirically" (see [88, 22]), but in the general setup the superstring perturbation theory is not mathematically well-defined. Moreover, these are not problems of the formalism: the same obstacles arise in less geometrical approach of Green and Schwarz [91].

3.6 Effective Actions for Background Fields

By analogy with the generating functionals for particles in external fields one may introduce the interaction of strings with background fields. The integration over the string degrees of freedom will give rise to certain effective functionals, depending already only upon the local fields in space-time. Such functionals are called the Fradkin-Tseytlin effective actions [57], and can be considered as the most efficient way for getting effective field theories from string theory.

Such an approach looks very transparent and clear from an ideological point of view. Indeed, at observable energies massive string modes are not excited and only the massless local fields "fly out" into our low-energy world. The interaction of string with local fields can be easily written down from certain symmetry requirements, say adding an exponential of the interaction term with the gauge field [18]

$$\int_{\partial\Sigma} dt \left(\dot{X}_M(t) A_M(X(t)) + \frac{e(t)}{2} F_{MN}(X(t)) \Psi_M(t) \Psi_N(t) \right) \tag{3.24}$$

(the ordered P-exponent in the non-Abelian case). The procedure here is the same as for relativistic particle, one should only remember that an integration in (3.24) is taken over the boundary of the world-sheet $\partial\Sigma$, while in the case of a particle the integration was taken along the whole world-line. This means that only the open strings interact with the vector fields. In the closed string sector the situation is similar, and the action is defined by the terms like (3.16), where the interaction (and thus the integration) is performed over the whole surface of the world-sheet.

In quadratic approximation the effective string actions *must* coincide with quadratic terms in the Lagrangians of the corresponding field theories for the background fields. The direct derivation of this correspondence is impossible due to vanishing of the two-point correlators on the world-sheets of the simplest topology (this is again a direct consequence of two-dimensional geometry). An indirect argument in favor of such a coincidence is self-consistency of the theory. Indeed, two-dimensional conformal invariance requires that background fields satisfy equations of motion, which in their turn would require the appropriate kinetic terms in the effective Lagrangians. The higher terms in background fields and derivatives in the effective actions follow straightforwardly from the calculation of string amplitudes.

One of the most interesting (and one of the few computable) examples of the non-local effective actions, arising for strings in the external gauge fields is the Dirac-Born-Infeld action (in any even-dimesnional space-time)

$$S_{DBI} = \int_{d^D x} \sqrt{\det_{MN}(G_{MN} + 2\pi\alpha' F_{MN})} \tag{3.25}$$

It comes out directly from the calculation of the effective string action for external electro-magnetic field, interacting with the string world-sheets of the open strings having the simplest possible topology of a disk [83].

This is a rather nontrivial fact – all the corrections in α', or loop corrections from the point of view of two-dimensional field theory (let us recall here that from the point of view of string theory any computation on disk counts only the "tree-level" contributions) sum up to the compact formula (3.25). This formula is really valid at large fields $F_{MN} \sim \alpha'^{-1}$ of the order of string tension. The action (3.25) has supersymmetric and even non-Abelian analogs which are rather interesting for the investigation of effective actions in nonperturbative string theory.

In the closed string sector one gets an effective action for the Einstein gravity

$$\int_{d^D x} \sqrt{G} e^{-2\Phi} \left(R(G) + \frac{1}{2}(\nabla\Phi)^2 + \dots \right) \tag{3.26}$$

[18] Notice that the operator (3.19) literally corresponds to the first term in formula (3.24), if one takes for the role of gauge field the solution to the equations of motion in the form of plane wave $A_M(X) \propto \epsilon_M \exp(ip \cdot X)$.

where $G \equiv \det_{MN} G_{MN}$, with the only difference that the scale or normalization of "string" metric differs from the "scale" or normalization of the Einstein metric by (exponent of the) vacuum value of the dilaton field Φ. It leads in particular to the fact that the "real" Newton constant or the Planck mass in ten-dimensional theory is connected to the string tension via

$$\gamma_{\mathrm{N}}^{(10)} = \left(M_{\mathrm{pl}}^{(10)}\right)^{-8} = g_{\mathrm{str}}^2 \alpha'^4 \tag{3.27}$$

where $g_{\mathrm{str}} = \langle \exp(\Phi) \rangle$. This relation will be essentially used below in discussion of the nonperturbative string theory.

4 Strings without Strings. Non-perturbative Theory

4.1 M-theory

Let us turn now to some achievements in string theory of the last ten years, related mostly with the attempts to go beyond the perturbation theory. As we already discussed in the context of quantum field theory, one immediately looses any "solid background" since this is the field where there is no reliable formalism. All possible statements can be based on a few "semi-qualitative" considerations [19]. Nevertheless, these attempts can have some success and there still exists a hope that they will be mostly successful in the framework of string theory. This hope is based on the existence of certain deeply "hidden" symmetries which may manifest themselves at nonperturbative level.

Note here that in contrast to the widely spread opinion about the pure mathematical character of the problems of string theory (which is not too far from being true if we restrict ourselves to the string perturbation theory), the problems of nonperturbative string theory have more fundamental and physical character. Let us repeat that the main problem is that nonperturbative string theory (as well as nonperturbative quantum field theory) does not exist in adequate physical form, i.e. does not exist in the form of any reasonable formalism. What is called at the moment nonperturbative string theory or M-theory is just a set of purely "philological" postulates reminding one, say, the "Butlerov theory", well-known from the high-school course of organic chemistry .

The main hypothesis formulated at present in this or that way implies existence of some unique nonperturbative string theory or M-theory [109, 110] (see also the reviews [19]-[21]) which has a large set of vacua understood in the sense of perturbative string theory. In other words, the perturbation theory around these vacua corresponds to (different!) two-dimensional conformal field theories considered above, interacting via anomaly with two-dimensional gravity. The fact that different perturbative expansions describe different phases of the same theory is encoded in the so called duality – not very well-defined and often only intuitively understood similarity of certain objects from the different phases of M-theory.

In the limiting case this means that there exist duality transformations, relating different quantities in quantum field theories. These relations can be established even between the quantities in absolutely different regimes, for example the particle-like states in one theory may be related to the soliton-like states in the dual one and vice versa. This is the reason why such duality cannot in practice be verified by standard methods of quantum field theory (except maybe in the two-dimensional theories, where, for example the well-known duality between the sine-Gordon and Thirring models exists). On the other hand it allows to consider the well-known problems from an absolutely new perspective and sometimes leads to surprising new results.

The hypothetical properties of M-theory make it a little bit similar to the field theory which contains together with "particle-like" states the collective nontrivial excitations like solitons, monopoles etc. However, in contrast to conventional quantum field theory, depending on the values of parameters or moduli of M-theory (for example the vacuum condensates of the scalar fields) the same observable objects (say electrically and/or magnetically charged particles) may be described equally as elementary and/or soliton-like particles with *different* field-theoretical Lagrangians.

Speaking about M-theory we will still use the term "string theory" despite the fact that in nonperturbative theory the very concept of fundamental one-dimensional extended objects acquires much more "hidden" form. In various considerations of M-theory a huge amount of hypersurfaces of arbitrary dimension (or, better to say,

[19] An exception can be found, if any, in the framework of so called "discretized" versions of quantum field theory, for example, in the so called "lattice" theories which are beyond the scope of this review. Note also that the progress in understanding of non-perturbative effects in lattice gauge theories is seriously "screened" by additional problems of the correspondence between the lattice theory and its continuum limit.

of arbitrary co-dimension) take part. From the naive point of view the one-dimensional extended objects are not at all singled out among other, and strings are just particular case of so called p-branes (number p measures the dimension of brane). For example, particle corresponds to $p = 0$, string – to $p = 1$, the membrane from which is derived the word *brane*), – to $p = 2$ and so forth.

However, the special role of strings is still caused by the fact that only strings can pretend to be the fundamental objects. We cannot really add anything here to the arguments of sect. 3.1, with the only difference being that now one should discuss separately the particular domains of moduli space of nonperturbative theory. In different domains there can exist (and do exist) different theories of fundamental strings. In such situation the fundamental string of one of perturbative theories can be, generally, the heavy "composite object" in another perturbative theory. Moreover, only strings are naturally charged with respect to vector fields which leads, on one hand, to the non-Abelian theories, and on the other hand the gauge invariance of the theories of vector fields (and gravity) allows opportunity for existence of *light* strings (more strictly light excitations of strings) while light membranes etc are absent.

The notion of duality, at least in the sense to be used below has mostly stringy origin and is related to the properties of complex manifolds often arising already in perturbative string theory. In perturbative string theory these properties belong to "unobservable" geometry of world-sheets, but, quite unexpectedly, analogous properties arise in the context of complex manifolds, being the "auxiliary" nontrivial part of the multidimensional space-time. The dualism between the structures on world-sheets and in target-space is rather surprising and not yet well-studied phenomenon in string theory, a manifestation of this intrinsic connection – the relation between world-sheet and space-time supersymmetries was already discussed in sect. 3.5. The simplest example of duality between anomaly free string models – the so called T-duality relating IIA and IIB superstring theories – is a direct consequence of the famous $R \leftrightarrow \frac{\alpha'}{R}$ duality, to be considered in detail in sect. 4.2. Other duality transformations typically relate to each other two theories with at least one of them being in strong coupling phase. Thus, their verification is an absolutely nontrivial problem.

Let us now try to list the main *postulates* of M-theory:

- **M-theory and eleven-dimensional supergravity**. The low-energy limit of M-theory is supergravity in a space-time of $D = 11$ dimensions [20] [72]. This is the maximal possible supergravity and, thus, maybe the only distinguished and nice theory from all supergravity models. Its bosonic sector contains only the metric G_{MN} and antisymmetric tensor field (the 3-form) C_{MNK}. The only (dimensional) parameter in this theory – the eleven-dimensional Planck mass $M_{\rm pl} \equiv M_{\rm pl}^{(11)}$. Under dimensional reduction of eleven-dimensional supergravity one gets the ten-dimensional supergravity of the type IIA – the field theory limit of IIA string theory. This leads to the relation between the square of string length (or inverse string tension) α', radius of the compact dimension R and eleven-dimensional Planck mass, which reads

$$\alpha' R M_{\rm pl}^3 = 1 \tag{4.1}$$

The relation between ten-dimensional and eleven-dimensional Planck masses

$$M_{\rm pl}^9 R = \left(M_{\rm pl}^{(10)}\right)^8 = \frac{1}{g_{\rm str}^2 \alpha'^4} \tag{4.2}$$

can be obtained directly from the reduction of the Einstein action of supergravity (the first equality in formula (4.2)). The connection between ten-dimensional Planck mass and string coupling constant (the second equality in (4.2)) is a consequence of the difference between the string and gravitational "definitions" of metric, differing by $\langle \exp(-2\Phi) \rangle = 1/g_{\rm str}^2$, where Φ is the dilaton field (see (3.26), (3.27)). Altogether this leads to the equality

$$R = g_{\rm str}\sqrt{\alpha'} \tag{4.3}$$

demonstrating that with the growth of the string coupling constant $g_{\rm str}$ the radius of a hidden compact dimension R *blows up*. This leads to a possible interpretation of M-theory as a string theory in a strong coupling regime.

- **M-theory as type IIA string theory at strong coupling**. M-theory is *not* a theory of fundamental strings in the sense of sect. 3.3, already because there are no anomaly free perturbative string theories

[20] Dimension $D = 11$ is singled out (by a slightly strained arguments) already directly from geometric interpretation of the Standard Model with the gauge group $U(1) \times SU(2) \times SU(3)$ (see, for example [1], page 275). If we consider that the group of Standard Model naturally acts on some manifold of compactification, then the natural dimension of such manifold can be determined as a sum of unity for the $U(1)$-factor, two ($\dim(\mathbf{S}^2) = \dim(\mathbf{CP}^1) = 2$) for the $SU(2)$-factor and four ($\dim(\mathbf{CP}^2) = 4$, if it is implied that group acts on complex manifold) for the $SU(3)$-factor. Together with four "visible" dimensions this gives $D = 1 + 2 + 4 + 4 = 11$.

with the space-time dimension $D = 11$. Nevertheless, the arguments presented above allow one to consider M-theory as a type IIA string theory at strong coupling, where the extra compact dimension shows up and the size of this dimension is related to the strong coupling constant by eq. (4.3).

- **Strings and extended objects in M-theory**. The analysis of extended objects being solutions to the equations of motion in M-theory (in reality – the equations of motion of eleven-dimensional supergravity) and their dimensional reduction to $D = 10$ leads to rather natural parallels between branes in M-theory and branes in string theory. Say, the hypothetical membrane of M-theory winding along the compact dimension becomes a string. One more similar relation will be discussed below in sect. 4.4 when we are going to discuss the exact nonperturbative results in supersymmetric gauge theories. It turns out that it is M-theory's 5-brane which plays the main role in geometric formulation of these results.

As well as ten-dimensional perturbative string theory, the eleven-dimensional M-theory may manifest itself in a four-dimensional world only after some "compactification". One of the differences between perturbative and non-perturbative theories in this context is that presence of the extended objects leads after compactification to some new nontrivial effects. A remarkable property of supersymmetry is its relation to the complex geometry of (especially nontrivial part of) space-time. It is reflected in the fact that the nontrivial complex manifolds of string compactification correspond to effective supersymmetric quantum field theories. Parameters of such theories (coupling constants, vacuum condensates, masses etc) are parameters or moduli of the complex manifolds of the corresponding string compactification, for example of the Calabi-Yau manifolds [2]. The duality transformations in this case can be identified with action of the corresponding modular group.

In order to get the macroscopic four-dimensional gauge theory, one should find some four-dimensional reduction. There is a standard way in string theory coming back to the old Kaluza-Klein idea: the full space time can be presented as a direct product of four-dimensional Euclidean space and some complex manifold K. The "internal" space K determines the "color" properties of the theory, the number of four-dimensional supercharges etc. Supersymmetry requires the compact manifold K to be the three-dimensional complex manifold in the ten-dimensional picture (or, say, to be the product of the three-dimensional complex manifold with a circle from the eleven-dimensional point of view).

Moreover, it turns out that sometimes the nontrivial part of this three-dimensional complex (or six-dimensional real) manifold can be presented by a one-dimensional complex curve (or two-dimensional Riemann surface Σ). Starting from eleven-dimensional M-theory one should choose some particular compactification scheme down to four dimensions, such that the resulting theory would get an appropriate four-dimensional supersymmetry, the required gauge group (in majority of real situations $SU(N)$) and an appropriate set of matter multiplets. According to [116], there exists a compactification scenario when the complex geometry can be formulated in terms of Riemann surfaces and this scenario leads exactly to the Seiberg-Witten effective theories [73].

It is the (complex) analytic structure which distinguishes a class of theories where the exact nonperturbative results can be formulated. These results are formulated using the technique of holomorphic (meromorphic) functions. The idea to use holomorphic objects goes back to the application of complex analysis to the theory of instantons and the Belavin-Knizhnik theorem [58, 9] of perturbative string theory, see sect. 3.3. In the simplest class of problems under discussion the moduli of physical theories may be identified exactly with the moduli of *one-dimensional* complex manifolds – the (space-time!) complex curves or Riemann surfaces Σ, which *a priori* have no relation to the world-sheets of string theory. However, to study these objects one may successfully use the same technical tools which were used when studying the perturbative string theory (see sect. 3). An analogous picture may be expected for the theories where physical moduli spaces are identified with the moduli spaces of higher-dimensional complex manifolds (two-dimensional complex manifolds $K3$, Calabi-Yau three-folds etc, see details, e.g. in [2]). Moreover, there exists a unifying picture of string compactification which implies that complex curves can be considered as degenerate cases of more general compactification manifolds, for example when the Calabi-Yau manifold effectively degenerates into one-dimensional complex curve Σ [112]. A nontrivial topological structure of the curve Σ is essentially nonperturbative information, since in the perturbation theory this curve arises only "locally" as a scale parameter. This means, in particular, that the string effects play an essential role in the structure of the exact nonperturbative solutions of gauge theories and the topological degrees of freedom, playing a decisive role for the construction of an effective theory, are directly related to "windings" of strings around nontrivial cycles in the manifolds of string compactifications.

4.2 Strings in Compact Dimensions

In the brightest form the difference between string theory and quantum field theory appears in the case of topologically non-trivial space-time, and the simplest example of such space-time is the space-time with compact dimensions or just a "box" with periodic boundary conditions at the ends. The structure of such "compactified" string theories implies the existence of a very nontrivial symmetry (duality) relating different string models. In particular, these models can be related in such a way that the perturbative regime in one of the models allows to propose some reasonable hypothesis about nonperturbative effects in another. In other words, duality transformations allow to consider string models as perturbative expansions (3.12) as expansions around different vacua of the same theory. The only weak point (at present) of this concept is the absence of any reliable or strict statements in the mathematical sense.

The main example of duality is symmetry in the theory of *closed* strings in a space-time with compact dimensions (in the simplest case – with the only co-ordinate taking values on some "circle" $\phi \sim \phi + 2\pi R n$, with $n \in \mathbf{Z}$ being any integer). The spectrum of such theory and one-loop partition function are invariant under discrete transformation $R \leftrightarrow \frac{\alpha'}{R}$ [92]. This invariance follows from the fact that in addition to the standard discrete spectrum of particle on a circle with the quantized momentum $p \propto \frac{n}{R}$, $n \in \mathbf{Z}$ (existing certainly as well in ordinary quantum field theory with compact dimensions) there exists also another type of string excitations: a string can wind around a circle and the energy of such winding mode is $\frac{mR}{\alpha'}$, also with $m \in \mathbf{Z}$.

In the "decompactification" limit $R \to \infty$, the first part of spectrum will become continuous (again, as in ordinary quantum field theory), while the string winding excitations would become infinitely heavy and their contribution to the partition function can be neglected. However, the full spectrum

$$M_{n,m}^2 = \left(\frac{n}{R}\right)^2 + \left(\frac{mR}{\alpha'}\right)^2 \qquad \forall\, n, m \tag{4.4}$$

is obviously *invariant* under the change $R \leftrightarrow \frac{\alpha'}{R}$.

The presence of the second term, or the spectrum of string winding modes in eq. (4.4) is sometimes interpreted as stringy modification of the uncertainty principle. Indeed, expression (4.4) allows to think that the uncertainty principle $\Delta X \sim \frac{1}{E}$ is valid literally up to scales of the order of $\sqrt{\alpha'}$, while beyond this scale the formula should rather be replaced by something like $\Delta X \sim \frac{1}{E} + \alpha' E$.

It is relatively easy to see that the duality transformation $R \to \frac{\alpha'}{R}$ leaves invariant the holomorphic quantities, say the current $\partial\phi_L(z) \to \partial\phi_L(z)$, but changes the sign of the anti-holomorpic ones: $\bar\partial\phi_R(\bar z) \to -\bar\partial\phi_R(\bar z)$. It means, for example, that the operators of emission and absorbtion of "particles" of the form

$$V_p \propto \exp\left(ip\phi(z,\bar z)\right) = \exp\left(ip\phi_L(z) + ip\phi_R(\bar z)\right) \tag{4.5}$$

become *non-local* (from the point of view of the field $\phi(z,\bar z) = \phi_L(z) + \phi_R(\bar z)$) operators of the world-sheet "vortices"

$$\exp\left(ip\phi_L(z) - ip\phi_R(\bar z)\right) \tag{4.6}$$

and vice versa.

The same is true for the action of the duality transformations $R \to \frac{\alpha'}{R}$ on the holomorphic and/or anti-holomorphic (on the equations of motion) world-sheet fermions: $\Psi_L(z) \to \Psi_L(z)$, but, at the same time $\Psi_R(\bar z) \to -\Psi_R(\bar z)$. This immediately leads to nontrivial consequences for the type II superstrings in ten-dimensional space-time $\mathbf{R}^9 \times \mathbf{S}^1$ with one compact dimension. One can forget for a moment about the nine non-compact co-ordinates and consider what happens in such a theory under the transformation $R \to \frac{\alpha'}{R}$.

In the bosonic sector the winding modes still replace the Kaluza-Klein modes and vice versa, but the components of the two-dimensional fermionic fields Ψ along the compact direction corresponding to the left- and right- movers behave differently: one preserves the sign while the other one changes it. It follows then that the "Γ_5"-matrix, and therefore the operator of chirality projection changes sign only in one of the sectors. Hence, the non-chiral IIA theory under the transformation $R \to \frac{\alpha'}{R}$ turns into the chiral IIB theory and vice versa. The transformation $R \to \frac{\alpha'}{R}$ in multidimensional space-time with a single compact direction, exchanging the type IIA and type IIB theories is usually called T-duality. This is the only duality of string theory which can really be verified, since it relates the theories, which can be both considered at weak coupling. In a similar way T-duality relates the heterotic string models with the gauge groups $SO(32)$ and $E_8 \times E_8$.

Now, if we consider an effective action for string theory, say, in $D+1$ dimensions and reduce it to D dimensions, the size of the compact dimension arises as factor in front of the (D-dimensional) action, and can be further interpreted as a coupling constant. It allows one to turn $R \leftrightarrow \frac{\alpha'}{R}$-duality into relation between the effective theories such that one of these theories is at strong coupling while the other is weakly coupled. As a result of such reasoning one gets a *hypothesis* that some quantum field theory on a given manifold and at weak coupling is equivalent [21] to a different theory, generally on a different manifold and in the strong-coupling regime. It is quite surprising that applying this sort of arguments to particular supersymmetric gauge theories, it is possible sometimes to make explicit predictions about the exact spectra and exact form of low-energy effective actions.

To finish this section let us stop once more at so called "mirror symmetry" in string theory [23, 24, 25, 26]. We are not going to discuss the mathematical issues of this problem, related to the fact that string theory allows one hypothetically to establish certain relations between the complex and Kähler structures of some manifolds. For us it is more important that string theory in principle

[21] In the above sense. Such equivalence usually implies (partial) coincidence of spectra and *certain* correlation functions in dual theories.

possesses the possibility of "non-distinction" of the space time, in the sense that for given string model the space-time may not be determined uniquely. The simplest example of such phenomenon is discussed above – string models on the circles with the radii R and α'/R coincide at least at the level of spectrum [22]. Passing from circles to tori it is easy to see that the same symmetry is preserved. Under such process the type A theory on two-dimensional torus $\mathbf{T} = \mathbf{S}^1_{R_1} \times \mathbf{S}^1_{R_2}$ would become equivalent to the type B theory on the torus $\tilde{\mathbf{T}} = \mathbf{S}^1_{R_1} \times \mathbf{S}^1_{1/R_2}$ and vice versa. Notice now that the area of the torus Area($\mathbf{T}$) $= R_1 R_2$ and the modulus of complex structure $\tau(\mathbf{T}) = iR_1/R_2$ are up to imaginary unity, in different order, correspondingly the modulus $\tau(\tilde{\mathbf{T}})$ and area Area($\tilde{\mathbf{T}}$) of the "mirror torus" $\tilde{\mathbf{T}}$. Thus, we come to the statement of "mirror symmetry" about the equivalence of the A and B theories on mirror manifolds – the manifolds for which the moduli of complex and Kähler structures replace each other.

The physical nature of the mirror symmetry is rather transparent, though it contains a paradox at first glance. Replacement of momentum by the energy of the winding mode roughly speaking corresponds to the replacement of momentum by co-ordinate, and therefore the mirror symmetry is in some sense the symmetry between co-ordinates and momenta. It is clear that our world does not have such a symmetry, since we can always single out the space of co-ordinates or configuration space and the phase space is its cotangent bundle.

Hence, what should we do with mirror symmetry in string theory? The resolution of this puzzle is in the simple fact that such symmetry is possible only at the scales of order of $\sqrt{\alpha'}$, for example from dimensional requirement $p \leftrightarrow x/\alpha'$. Therefore, the mirror manifolds identified by string theory are in principle unobservable in the "macroworld"! Moreover, at such scales the phase space may not necessarily be a cotangent bundle. Say, the quantum mechanics of spin, formulated in adequate terms (see, for example, [93, 94]), corresponds to the phase space, having configuration of sphere, which is not at all a cotangent bundle. Another, maybe even more simple example from quantum mechanics is a particle in magnetic field. In this example there is a "natural replacement" of the configuration plane transversal to the direction of magnetic field by the "phase plane" on the distances of the order of magnetic length $l \sim \sqrt{\hbar c/eB}$.

4.3 Dimensional Reduction in String Theory and D-branes

Formula (4.4) leads to rather nontrivial conclusions about dimensional reduction in string theory. In field theory or the theory of pointlike particles the second term on the right hand size of (4.4), proportional to $(\alpha')^{-2}$, can be omitted and we obtain the conventional Kaluza-Klein spectrum. For the compactified quantum field theory it means that reducing the field theory from D to $(D-1)$ dimensions via the compactification of one dimension onto the circle of radius R with further limit $R \to 0$, the D-dimensional field can be conveniently written in terms of the Fourier series (not the Fourier integral) with respect to compact co-ordinate x_0

$$\phi(x, x_0) = \sum_n \exp\left(i\pi n \frac{x_0}{R}\right) \phi_n(x) \tag{4.7}$$

After substitution of this expansion into the action

$$\int_{d^{D-1}x} dx_0 \sum_{M=1}^{D} (\partial_M \phi)^2 = \int_{d^{D-1}x} \sum_n \left(\sum_{\mu=1}^{D-1} \partial_\mu \phi_n \partial_\mu \phi_{-n} + \frac{n^2}{R^2} \phi_n \phi_{-n} \right) \tag{4.8}$$

one gets the sum over $(D-1)$-dimensional fields $\phi_n(x)$ with the masses, exactly corresponding to the first term in (4.4). At $R \to 0$ all fields with $n \neq 0$ become infinitely heavy and at distances much more than R one may forget about them. Thus, after compactification and dimensional reduction we obtain from D-dimensional field theory the field theory in $(D-1)$ dimensions.

This rather natural conclusion remains intact even in the case of the open string theory, where the nontrivial winding modes corresponding to the second term in eq. (4.4) are absent. However, for the theory of a closed string one comes to a different conclusion. In the limit $R \to 0$ the Kaluza-Klein modes with the masses n/R still would become infinitely heavy, i.e. inessential for the limiting spectrum, but, in contrast to them, masses of all states corresponding to windings vanish! This means that at such reduction from D dimensions to $(D-1)$ dimensions, the Kaluza-Klein "tower" corresponding to an extra dimension disappears as in field theory ... but in the same procedure the equivalent "tower of fields" reappears due to the light at $R \to 0$ modes of the closed string winding around the compact direction. Thus, the extra tower of fields remains in the spectrum of closed string, i.e. no reduction to $(D-1)$ dimensions really happened and the theory remains D-dimensional!

Now, consider the same procedure in the theory with both closed and open strings. The conclusion is a bit of paradox: as $R \to 0$ closed strings would be still propagating in D-dimensional space-time, while the theory of open strings will be $(D-1)$-dimensional. Alternatively, if we require consistency and "smooth" behavior of string theory under the change of parameter or moduli R – the size of a compactified dimension, one has to allow the existence of absolutely new nontrivial vacua, containing certain distinguished hypersurfaces (the example

[22] Let us recall once more that identifying different string models by duality transformations one should strictly fix what is exactly identified and in what sense. Typically only the spectrum and *some* correlation functions are borne in mind.

considered above contained a hypersurface of unit codimension, however, it is easy to see that compactifying several dimensions the codimension can be made arbitrary). These hypersurfaces are characterized by the fact that only there the open strings can keep their ends. In modern terminology such hypersurfaces are called the Dirichlet or D-branes, and the volume between branes is called the bulk.

Let us now list the main properties of D-branes, essential for the study of nonperturbative string theory:

- Since vector fields arise in the open string sector (see sect. 3.1), in the theory (or, better to say in the vacuum) with D-branes the vector fields are *localized* on the D-brane's hypersurfaces. Hence, D-branes proposed a new, purely string mechanism of the localization of vector fields, which is absent in quantum field theory. Notice also, that the theory with open strings in all D-dimensional space-time can be interpreted as a vacuum with the Dirichlet brane (or several Dirichlet branes in the case of nontrivial Chan-Paton factors) of dimension $p = D - 1$, see fig. 9 and sect. 4.4).

- In the theories with space-time supersymmetry D-branes are the BPS states, invariant under the action of half of the supersymmetry generators. This is due the fact that in the open string sector there are twice fewer supersymmetry generators than in the closed sector, since the fields on the boundary of the world-sheet are constrained by the boundary conditions. The BPS nature of D-branes is also related directly to the fact that they are charged with respect to antisymmetric tensor fields of the Ramond-Ramond sector. Namely, the Dp-brane is charged with respect to the $(p+1)$-form, which can be integrated over the world-volume of the Dp-brane as $\int C^{(p+1)}$, and the corresponding charge arises as a central extension of the supersymmetry algebra. This central extension breaks, however, the D-dimensional Lorentz-invariance as well as the very existence of the hypersurface of D-brane.

- The D-brane tension is proportional to the *first* power of the string coupling constant. One of the arguments supporting this relation is interaction of D-brane with the *open* strings, whose perturbation theory contains the expansion in $g_{\rm str}$, and not in $g^2_{\rm str}$, see sect. 3.3. This distinguishes D-branes from so called solitonic branes, interacting only with closed strings. The corresponding effective action of the background fields (see sect. 3.6) can be roughly written as

$$\int_{d^D x} \sqrt{G}\left(e^{-2\Phi}(R(G) - H^2) - (dC)^2\right) \tag{4.9}$$

where Φ is the dilaton, $\langle \exp(-2\Phi)\rangle = g^{-2}_{\rm str}$; $R(G)$ is the curvature of D-dimensional metric, $G \equiv \det_{MN} G_{MN}$, $H = dB$ is the field-strength of antisymmetric tensor field, related to the solitonic branes while dC is the field-strength of the Ramond-Ramond $(p+1)$-forms. It is the different dependence on dilaton of the terms $(dB)^2$ and $(dC)^2$ in eq. (4.9) that leads to the fact that the "thickness" of the solitonic brane does *not* depend on $g_{\rm str}$ (for constant dilaton equations obtained from variation of the terms $\sqrt{G}(R(G) - H^2)$ in formula (4.9) and their solutions do not depend on $g_{\rm str}$), and its mass or tension is proportional to $g^{-2}_{\rm str}$, while the "thickness" of the D-brane (solution to the equations following from variation of the terms $\sqrt{G}\left(e^{-2\Phi}R(G) - (dC)^2\right)$ in (4.9)) is proportional to $g_{\rm str}$, and its tension is proportional to $g^{-1}_{\rm str}$. This means that at weak coupling D-brane can indeed be considered as a very thin hypersurface "glued" to the ends of the open strings.

Note, that due to the absence of "normal" nonperturbative theory these properties are established only with the help of certain mostly qualitative arguments (see, for example, [3, 17]). In what follows we will restrict ourselves to a "minimal use" of these properties, i.e. we will use them only where the D-brane picture leads to more or less clear physical consequences.

4.4 D-branes and non-Abelian Gauge Fields

Let us now discuss in detail how the (four-dimensional) supersymmetric gauge theories arise in the context of string theory. One should start with any supersymmetric string theory without anomalies. There exist several examples of such theories (defined originally as *perturbative* expansions in terms of the path integrals (3.12)) and their basic feature is that they live in $D = 10$ and have at least $\mathcal{N} = 1$ ten-dimensional space-time supersymmetry (see the end of sect. 3.5).

One of the main ingredients of the relation between strings and gauge theories are the above mentioned appearence of the D-brane configurations in non-perturbative string theory [61, 113]. D-branes are classical ("heavy") objects which can be thought of as certain hypersurfaces in a target space and whose basic feature is the possibility of interaction via emission and absorption of open strings (see fig. 9) – even in the theories with

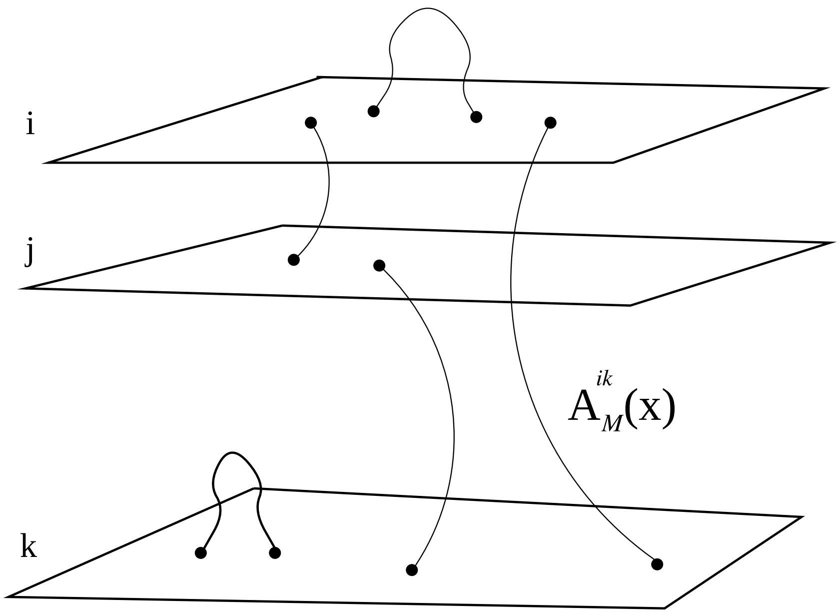

Figure 9: *D-branes. The interaction is carried by strings attached by their ends to different D-branes or parts of the same D-brane. In the background of several D-branes one naturally gets non-Abelian vector fields in the spectrum of strings since the fields become labeled by the numbers of D-branes they are attached to.*

no bulk open string interactions (for simplicity we will restrict ourselves only to such theories, called as type II theories, see sect. 3.5). As we already discussed in sect. 4.3, such hypersurfaces naturally arise in compactified string theory, implying that it behaves "smoothly" under the change of parameters of the compact manifold.

It is easy to see that the configuration of N parallel D-branes on fig. 9 leads naturally to the $SU(N)$ gauge group (more strictly to the group $U(N) = SU(N) \otimes U(1)$ with inessential for the fields in the adjoint representation $U(1)$ factor), broken down generally to $U(1)^{N-1}$. Indeed, consider N parallel D-branes, then the (oriented) open string stretched between the i-th and j-th brane ($i, j = 1, \dots, N$) (see fig. 9) contains a vector field $\mathbf{A}^{ij}$ in its spectrum. The mass of this vector field is proportional to the length of the string (since the energy or mass of a string is proportional to its length), i.e. to the distance between the i-th and j-th branes.

Thus, the $U(1)^{N-1}$ massless gauge fields will come out of the strings with both ends glued to the same D-brane, while the fields $\mathbf{A}^{ij}$ with $i \neq j$ will acquire the "Higgs" masses, proportional to the vacuum condensates of scalar fields (more strictly to the differences of these condensates for the corresponding components). These vacuum values are determined by the "transverse" co-ordinates of the D-brane $\phi \sim \frac{\sqrt{\vec{x}_\perp^2}}{\alpha'}$. Thus if the open strings themselves naturally lead to the appearance of massless vector gauge fields, the open strings in D-brane vacua rather naturally correspond to the theories with (in general broken) non-Abelian gauge symmetry [23].

The next step is – again looking at fig. 9 – to see how from ten-dimensional string theory one gets for such a configuration a theory in a much fewer number of dimensions (an ideal result would be to get four-dimensional theory). Indeed, it is easy to understand that the gauge theory "localizes" to the D-brane world-volume, i.e. the real number of vector indices is equal to the dimension of this world-volume. The D-brane hypersurface breaks full ten-dimensional Lorentz-invariance, therefore only the components corresponding to the directions "along" the world-volume form a real vector. The rest of the components, from the point of view of unbroken space-time theory on the D-brane world volume look like set of scalars, what is in complete analogy with the dimensional reduction of the theory of a vector field (see, for example [42]).

The Dirichlet p-brane world-volume [24] has dimension $p+1$ (including time!), i.e. naively in order to get four-dimensional gauge theory one should consider parallel D3-branes. This scenario is quite possible but gives rise to $\mathcal{N} = 4$ supersymmetry in four dimensions; in order to get less trivial $\mathcal{N} = 2$ (or even $\mathcal{N} = 1$) theory it is better

[23] Let us recall that before this fact was understood, non-Abelian gauge theories were constructed "by hand", "gluing" quarks to the ends of open strings (see fig. 3), or introducing the non-Abelian Chan-Paton factors [74] directly into string amplitudes.

[24] To avoid misunderstanding let us again point out the accepted terminology. D-brane is short for "Dirichlet brane" and has no relation with the dimension of this hypersurface, which is conventionally noted by the letter p. Sometimes even the notation Dp-brane is used, i.e. the p-dimensional Dirichlet brane with the world-volume of dimension $(p+1)$. Let us repeat once more that $p = 2$ corresponds to a membrane (the origin of the word "brane"), one would often meet in the literature, D1-branes, or D-strings, D0-branes or Dirichlet particles or even D(-1)-branes or D-instantons, as well as branes of dimensions $2 < p \leq D - 1$, where in the last inequality D means already the dimension of space-time and does not come from the word Dirichlet.

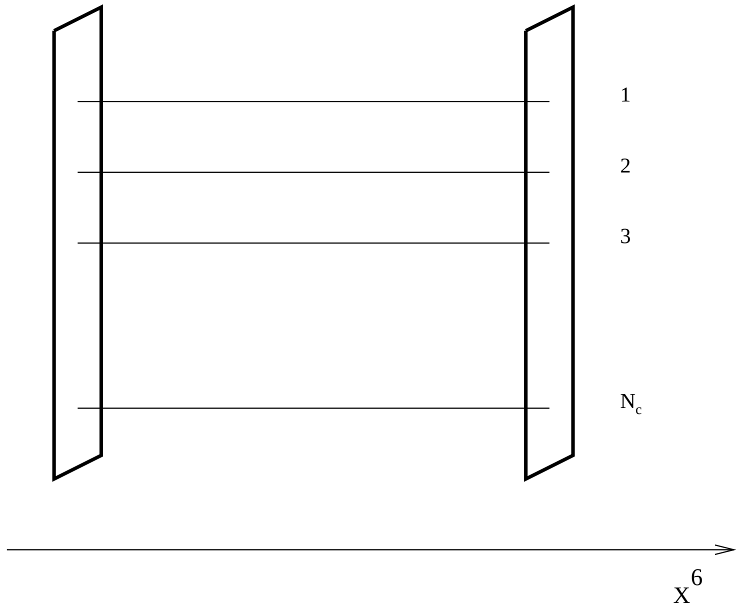

Figure 10: 4-branes restricted by 5-branes to a finite volume (in the horizontal x^6-direction) give rise to macroscopically 4-dimensional theory.

to use another option, the Diaconescu-Hanany-Witten "ladder" brane configuration [114, 116] with N parallel D4-branes stretched between two vertical walls (see fig. 10), so that naive five-dimensional D4 world-volume theory becomes macroscopically (in the light sector) four-dimensional by the famous Kaluza-Klein argument for a system compactified on a circle or put into a box. Certainly there are many other constructions based on discrete symmetries, orientifolds etc, however the "brane zoology" is beyond the scope of this review (see, for example [30]) and we will discuss only the simplest "ladder" example, especially since it is this example that corresponds to one of the strongest statements about non-perturbative supersymmetric gauge theories.

The role of vertical walls should be, best of all, played by 5-branes [116], then dimensional arguments lead to the logarithmic behavior of the macroscopic coupling constant $\frac{1}{g^2} \sim \log \mu$ (cf. with formula (2.4)). In the leading approximation this comes up since the corresponding "compact" co-ordinate, which turns into a coefficient in front of the action (2.2), has logarithmic behavior as a function of "transverse" directions, i.e. satisfies the *two*-dimensional Laplace equation, where the effective two dimensions are formed by the ends of D4-branes in 5-branes. More generally the fact that the logarithm (of the complex argument) is the Green function of the two-dimensional Laplace operator is one of the "foundations" for the D-brane constructions of supersymmetric gauge theories.

This picture of 4- and 5- branes in ten dimensions is certainly very rough and true only in a (quasi)-classical approximation. In particular it is naively singular at the points where 4-branes meet 5-branes. These singularities were resolved in a nice way in [116] where it was proposed to "raise" the whole picture into an eleven-dimensional target space of M-theory and to consider D4-branes as M-theory 5-branes compactified onto eleventh dimension with x^{10} being the corresponding extra compact co-ordinate. Then the picture in fig. 10 turns into the surface of a "swedish ladder" and apart from macroscopic directions $x^0, \ldots, x^3$ looks like a (non-compact) Riemann surface with rather special properties (see Fig. 11).

In other words, as a result of "resolution of singularities one gets a unique smooth 5-brane, which leaving aside four flat dimensions (x^0, x^1, x^2, x^3) looks like N cylinders $R \times S^1$ embedded into the target space along, say, (x^6, x^{10}) dimensions (and which can be parameterized by complex co-ordinate $z = x^6 + ix^{10}$). The cylinders are separated in the "orthogonal" space $V^\perp = (x^4, x^5, x^7, x^8)$, but they are all glued together (see fig. 11) by vertical walls, and the "effective" two-dimensional subspace of $V^\perp$ can be described by the complex coordinate $\lambda = x^4 + ix^5$.

Let us try to establish the relation between the brane configurations and complex manifolds. The simplest way to describe a nontrivial complex manifold is analytic, i.e. by certain (polynomial) equations in multidimensional complex space $\mathbf{C}^n$. Let us demonstrate now how the pictures in fig. 9 and fig. 11 can be rewritten in terms of algebraic equations on complex variables.

Introducing the co-ordinate $w = \exp(z)$ to describe a cylinder, we see that the system of non-interacting branes (fig. 9) is given

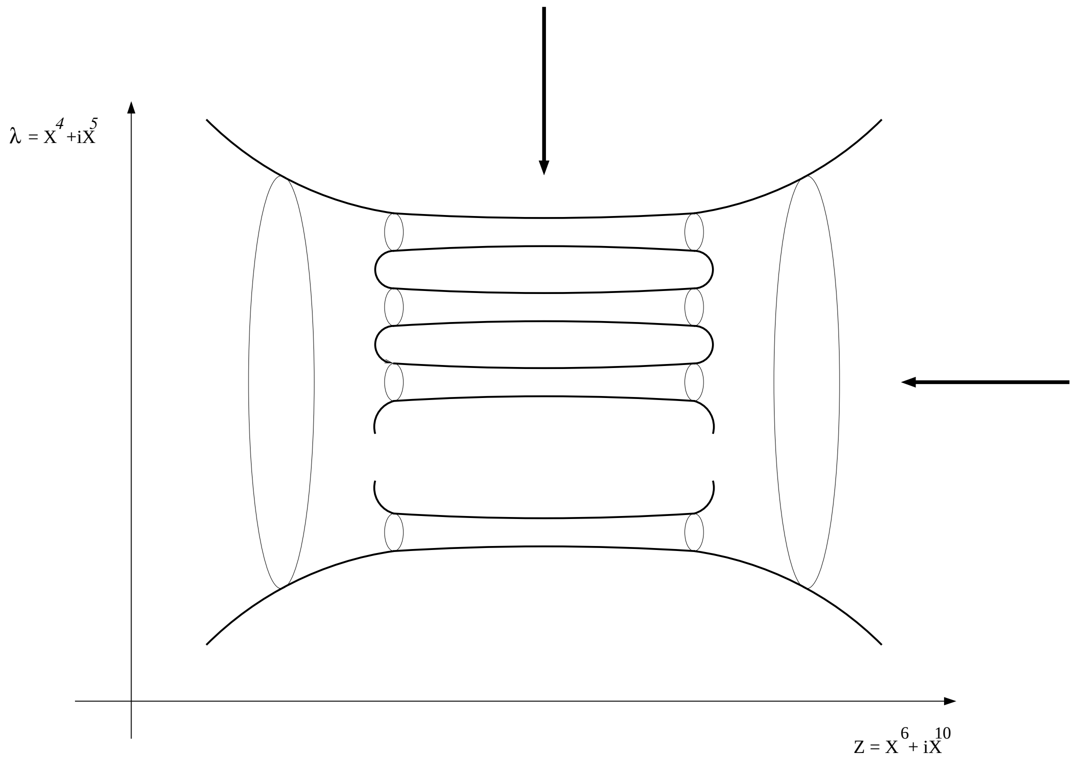

Figure 11: Brane configuration, represented as a result of "resolution" the previous picture – the "thin" ladder turns into a "swedish ladder" – the hyperelliptic Riemann surface being at the same time N-fold covering of the horizontal cylinder.

by the z-independent equation

$$P_N(\lambda) = \prod_{i=1}^{N}(\lambda - \phi_i) = 0, \tag{4.10}$$

while their bound state (fig. 11) is described by a complex curve Σ (a single equation on two complex variables)

$$w + \frac{\Lambda^{2N}}{w} = P_N(\lambda) \tag{4.11}$$

In the weak-coupling limit $\Lambda \to 0$ (i.e. $1/g^2 \sim \log(1/\Lambda) \to \infty$) one comes back to the set of disjoint branes (4.10). Eq. (4.11) presents an analytic formulation of fig. 11 – 5-brane of topology $R^3 \times \Sigma$ embedded *holomorphically* into a subspace $R^5 \times S^1$ (say, spanned by $x^1, ..., x^6, x^{10}$) of the full space-time.

A somewhat more transparent way to get the same equations is related to the theory of integrable systems [119] and uses the fact that in vacuum state the scalar fields satisfy the BPS-like condition – the first-order equation (cf. with (2.13))

$$D_M\Phi \equiv \partial_M\Phi + [A_M, \Phi] = 0 \qquad F_{MN} = 0 \tag{4.12}$$

It acquires exactly the form of eq. (4.12) when only one of the fields $\Phi^{(4)}, \dots, \Phi^{(8)}$ is nonvanishing – otherwise it would also contain the scalar interaction terms. This is essentially the case of the configuration depicted in fig. 11, which implies that some scalar field, say $\Phi \equiv \Phi^{(4)} + i\Phi^{(5)}$, develops a nonvanishing z-dependent vacuum expectation value. In order to explain or "derive" fig. 11, it is necessary to demonstrate that eq. (4.12) has a *non-trivial* solution $\Phi(z) \neq const$ and the reason for this is that non-trivial boundary conditions are imposed on Φ at $z \to \pm\infty$. This procedure is considered in detail in [119] and results in the so called *Lax representations* for the algebraic equations of nontrivial complex manifolds – in this case for the complex curves [120]. Under such procedure eq. (4.12) turns, for example, into

$$\bar\partial\Phi^{ij} + (q_i - q_j)\Phi^{ij} = m(1 - \delta^{ij})\delta(z - z_0) \tag{4.13}$$

with the solution

$$\Phi^{ij}(z) = p_i\delta^{ij} + m(1-\delta^{ij})\frac{\theta_*(z - z_0 + \frac{\mathrm{Im}\tau}{\pi}(q_i - q_j))\theta_*'(0)}{\theta_*(z - z_0)\theta_*(\frac{\mathrm{Im}\tau}{\pi}(q_i - q_j))}\exp\left((q_i - q_j)(z - \bar z)\right) \tag{4.14}$$

where $\theta_*(z)$ is the odd Jacobi theta-function. Equation $\det(\lambda - \Phi(z)) = 0$ (literally corresponding to the theory with broken $\mathcal{N} = 4$ supersymmetry) in the limit $m \to \infty$ and $\tau \to +i\infty$ with $m^N \exp(i\pi\tau) = \Lambda^N$ turns exactly into eq. (4.11), the details and references can be found in [8, 27].

In this way one can derive the analytic representation of the complex curve (4.11) "from first principles". The next step is to derive the effective action of the low-energy four-dimensional theory. According to [116], this problem can be solved starting from the effective action on the 5-brane world-volume or the theory of self-dual two-form $C = \{C_{MN}\}$, $dC =^* dC$. Roughly speaking it means that instead of open strings, as in fig. 9, the interaction is effectively performed by "open membranes". The theory of two-forms is essentially Abelian. Even if one introduces the matrices C^{ij}_{MN} in the adjoint representation of $SU(N)$ associated with the membranes attached between i-th and j-th cylinders, the non-Abelian interacting theory cannot arise since such interaction is inconsistent with the gauge invariance. Such a theory may contain only a non-linear interaction of non-minimal type – like $\mathrm{Tr}(dC)^4$, i.e. depending upon the tension of C. These terms, however, contain higher derivatives (powers of momentum) and they are irrelevant in the low-energy effective actions. The "Abelian" nature of the theory of two-forms makes the description of the Lax operator (vacuum expectation value of the scalars of the supermultiplet which describe the transverse fluctuations of the 5-brane), and thus the derivation of the shape of the curve Σ in the type IIA picture, a nontrivial problem. Instead, exactly due to the fact that the action on (flat) world-volume is essentially quadratic

$$\int_{d^6x} |dC|^2 + \text{supersymmetric terms} \tag{4.15}$$

there are no corrections to the form of the effective four-dimensional action, once the curve Σ is given. It is enough to consider the dimensional reduction of (4.15) from six down to four dimensions [116], implying that the two-form C can be decomposed as

$$C_{\mu z} = \sum_{i=1}^{N-1} \left(A^i_\mu(x) d\omega_i(z) + \tilde{A}^i_\mu(x) d\bar{\omega}_i(\bar{z}) \right) \tag{4.16}$$

where $d\omega_i$ are canonical holomorphic one-differentials on Σ, $d\bar{\omega}_i$ – their complex conjugate, and the fields A^i_μ, $\bar{A}^i_\mu$ depend only on the four co-ordinates $x = \{x^0, x^1, x^2, x^3\}$.

Choosing the metric on Σ to be such that $*d\omega_i = -d\omega_i$, $*d\bar{\omega}_i = +d\bar{\omega}_i$, the self-duality of C implies that the one-forms A and $\bar{A}$ in (4.16) correspond to the anti-selfdual and selfdual components of the four-dimensional gauge field with the curvature (tension) $F = \{F_{\mu\nu}\}$:

$$\begin{aligned} dA^i &= F^i -^* F^i \\ d\tilde{A}^i &= F^i +^* F^i \end{aligned} \tag{4.17}$$

It remains to substitute this into (4.15) to get T_{ij} – the period matrix of Σ (which depends on the vacuum expectation values of the transverse scalar fields once the shape of the curve Σ or its embedding into the (x^4, x^5, x^6, x^{10})-space is already fixed). The result for the four-dimensional effective action reads

$$\int_{d^4x} \mathrm{Im} T_{ij} F^i_{\mu\nu} F^j_{\mu\nu} + \text{supersymmetric terms} \tag{4.18}$$

where effective couplings are expressed through (the imaginary part of) the period matrix $\mathrm{Im}T_{ij} = \int_\Sigma d\omega_i \wedge d\bar{\omega}_j$ of the auxiliary Riemann surface (4.11). The action (4.18) coincides with the result of the Seiberg-Witten theory [73], up to the topological θ-term, which can be restored by slightly more delicate operating with the action of a self-dual two-form.

4.5 Seiberg-Witten Theory

The construction of the exact nonperturbative effective actions for the low-energy $\mathcal{N} = 2$ supersymmetric gauge theories is called Seiberg-Witten theory [73]. The exact nonperturbative formulas [73] contain the information about the spectrum of the BPS excitations ("W-bosons" and monopoles, see sect. 2.3) and the Wilsonian effective action of the light fields (see, for example, [41, 86]).

As we already pointed out in sect. 2.4, supersymmetry leads to strong requirements on the form of the effective action. In the case of $\mathcal{N} = 1$ supersymmetry in four dimensions the "classical" form of the superpotential W is *not renormalized* (and this allows to study vacua of the theory – the critical points of the superpotential $dW = 0$) while the kinetic terms are governed by the Kähler metric or the Kähler potential. For the extended supersymmetry the situation is even more restrictive – there are no Abelian potential terms (and it means that instead of distinct vacuum "points" one gets the continuous set of vacua described by parametric families or *moduli*) and the effective action, say for the vector multiplets, can be written in terms of a single holomorphic function of several complex variables [106, 73] – a *prepotential*. In other words the geometry of moduli space is *special* Kähler.

Let us turn now to the Seiberg-Witten theory for the $\mathcal{N} = 2$ supersymmetric Yang-Mills theory without matter [25]. The scalar potential in $\mathcal{N} = 2$ supersymmetric gauge theory is essentially "non-Abelian" and has the form $V(\boldsymbol{\phi}) = \mathrm{Tr}[\boldsymbol{\phi}, \boldsymbol{\phi}^\dagger]^2$. Its minima after factorization over the gauge group correspond to the diagonal ($[\boldsymbol{\phi}, \boldsymbol{\phi}^\dagger] = 0$), and in the theory with the $SU(N)$ gauge group – to the traceless matrices (2.6). Due to spontaneous breaking of the gauge group this results (in general position) in the effective $\mathcal{N} = 2$ Abelian gauge theory with the effective Lagrangian $\mathcal{L}_{\text{eff}}(\Phi_i)$, which can be written, say, in terms of the superfields Φ_i, whose vacuum values $\langle\varphi_i\rangle = \phi_i$ coincide with the diagonal elements of (2.6). Therefore the function of complex variables $\mathcal{F}(a) = \mathcal{F}(\boldsymbol{\phi})|_{\sum \phi_i = 0}$ (where the independent variables a_i – in perturbation theory – can be chosen,

[25] In supersymmetric gauge theories one usually means by matter only the multiplets of the fermionic and scalar fields in the fundamental representations of the gauge group – the analogs of quarks in usual (non-supersymmetric) QCD. In the theories with extended supersymmetry there are also the fermions and scalars in the adjoint representation – the superpartners of the gauge fields.

for example, as $a_i = \phi_i - \phi_N$, $i = 1, \ldots, N-1$) indeed determines the Wilsonian effective action for the massless fields by means of the following substitution

$$\mathcal{L}_{\text{eff}} \propto \text{Im} \int d^4\vartheta \mathcal{F}(\phi_i \to \Phi_i) = \text{Im}\left(\frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j}\right) F^i_{\mu\nu} F^j_{\mu\nu} + \text{supersymmetric terms} \tag{4.19}$$

Notice immediately that the effective action (4.19) exactly coincides with (4.18), after identifying the matrix elements of the period matrix $T_{ij} = \frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j}$ with the second derivatives of prepotential.

In $\mathcal{N} = 2$ perturbation theory formula (4.19 can be checked by explicit computation of quantum corrections, which in conventional $\mathcal{N} = 2$ supersymmetric gauge theory are reduced to the one-loop diagram (see fig. 1). Integrating over momenta propagating along the loop one comes to the result

$$T_{1-\text{loop}} \propto \sum_{\text{masses}} \log \frac{(\text{mass})^2}{\Lambda^2} \tag{4.20}$$

where $\Lambda \equiv \Lambda_{QCD}$ is a scale parameter of the theory and the sum in (4.20) is taken over the masses of propagating fields in the loop. In the easiest form this result can be written in terms of the "Coleman-Weinberg" formula for the prepotential

$$\mathcal{F}_{1-\text{loop}} = \frac{1}{4} \sum_{\text{masses}} (\text{mass})^2 \log \frac{(\text{mass})^2}{\Lambda^2} \tag{4.21}$$

In pure supersymmetric Yang-Mills theory all masses in (4.21) are generated by the Higgs effect (2.9), so finally the perturbative result (4.21) acquires the form

$$\mathcal{F}_{\text{pert}} = \mathcal{F}_{1-\text{loop}} = \frac{1}{4} \text{Tr}\left(\phi^2 \log \frac{\phi^2}{\Lambda^2}\right) \tag{4.22}$$

The same computation can be performed in the general case: one should take the sum of the terms like (4.22) corresponding to the contribution of each multiplet; the trace for each term $\text{Tr} \equiv \text{Tr}_R$ should be taken in the corresponding representation and the sign of each contribution depends of the type of the multiplet (it is "+" for the vector and "−" for the hypermultiplet). As for the massive excitations, it turns out that at least the BPS massive spectrum

$$M \propto |\mathbf{n} \cdot \mathbf{a} + \mathbf{m} \cdot \mathbf{a}_D| \tag{4.23}$$

is related to the prepotential $\mathcal{F}$ by the formulas [73]

$$\mathbf{a}_D = \frac{\partial \mathcal{F}}{\partial \mathbf{a}} \tag{4.24}$$

The integer-valued vectors $\mathbf{n}$ and $\mathbf{m}$ in eq. (4.23) correspond respectively to "electric" and "magnetic" charges of "surviving" $U(1)^{N-1}$ gauge group.

With the instantonic contributions things are not so simple. The well-known part contains the generic structure of the effective action which implies that prepotential has an asymptotic expansion for large values of the condensates $\langle\Phi\rangle \gg \Lambda$

$$\mathcal{F} = \mathcal{F}_{\text{pert}} + \mathcal{F}_{\text{inst}} = \frac{1}{4} \sum_{\{I\}} a^2_{\{I\}} \log \frac{a^2_{\{I\}}}{\Lambda^2} + \sum_{\{I\}} a^2_{\{I\}} \sum_{k=1}^{\infty} \mathcal{F}_{\{I\},k} \left(\frac{\Lambda}{a_{\{I\}}}\right)^{2Nk} \tag{4.25}$$

with some unknown coefficients $\mathcal{F}_{\{I\},k}$, where the multiindex I corresponds to different components of the vector $\mathbf{a}$. The terms with fixed k in the r.h.s. of (4.25) corresponds to the sector with fixed instantonic number k in the $SU(N)$ Yang-Mills theory. For example, in the $SU(2)$ case the integral over the size of each instanton has the form $\int \frac{d\rho}{\rho^5}$ giving rise to the Λ^{4k} scale dependence for k instantons. However, *all* coefficients $\mathcal{F}_{\{I\},k}$ in principle cannot be computed by standard field-theoretical methods. Each of them can be written in the form of some integral over the (each time different) moduli space of an instanton configuration, therefore their "relative normalization" simply cannot be defined. On the other hand, such normalization can be fixed in some "natural way", and all performed instantonic calculations (mostly with the $SU(2)$ gauge group) confirm the Seiberg-Witten hypothesis.

According to the Seiberg-Witten hypothesis the BPS masses $\mathbf{a}$ and $\mathbf{a}_D$ can be expressed through the periods of a meromorphic differential dS on auxiliary *Riemann surface* Σ and depend on the vacuum expectation values of scalar fields, as upon certain co-ordinates on the moduli space of complex structures of Σ. In particular, in

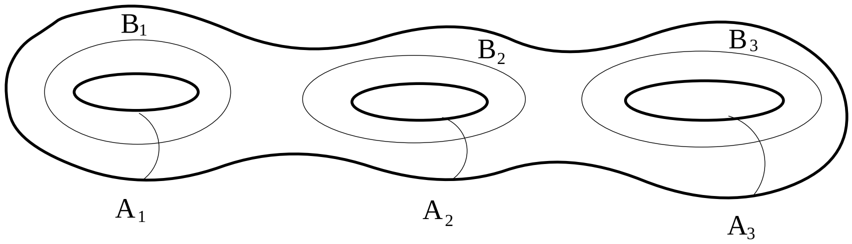

Figure 12: *Compact two-dimensional Riemann surface of genus $g = 3$. The canonical basis of* **A** *and* **B**-*cycles has the intersection form $A_i \circ B_j = \delta_{ij}$. An analogous picture arises in fig. 11 if one adds "by hand" both "infinity points" $\lambda = \infty$.*

these specific co-ordinates the matrix of effective charges $T_{ij}(\mathbf{a}) = \frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j}$ plays the role of the *period matrix* of Riemann surface Σ. For example, in the case of pure gauge theory with the $SU(N)$ gauge group the auxiliary Riemann surface has exactly the form (4.11) [107], where the coefficients of the polynomial $P_N(\lambda)$ are expressed through the vacuum values of the scalar fields (2.7). The exact quantum values of the BPS masses are related to the vacuum condensates through the *periods* over the so called **A**-cycles (see fig. 12)

$$\mathbf{a} = \oint_{\mathbf{A}} dS \tag{4.26}$$

for the W-bosons, and the **B**-cycles for the monopoles

$$\mathbf{a}^D = \oint_{\mathbf{B}} dS \tag{4.27}$$

of the meromorphic differential

$$dS = \lambda \frac{dw}{w} \tag{4.28}$$

whose properties ensure (see the details, say, in [8, 27]), that the period matrix of the Riemann surface (4.11) can be expressed in terms of the derivatives

$$T_{ij} = \frac{\partial a_i^D}{\partial a_j} = \frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j} \tag{4.29}$$

Eq. (4.11) can be explained (but not derived!) in the following way. Perturbatively, the masses of "particles" – the W-bosons and their superpartners are proportional to the differences of ϕ_i's or the *roots* of the "generating" polynomial (2.7). Thus they can be "extracted" from the polynomial (2.7) via the residue formula

$$m_{ij} \propto \oint_{C_{ij}} \lambda d \log P_N(\lambda) \tag{4.30}$$

which for a particular contour C_{ij} – a "figure-of-eight", drawn around the points $\lambda = \phi_i$ and $\lambda = \phi_j$ (see fig. 13) – gives rise directly to (2.9). The contour integral (4.30) can be viewed as defined on *degenerate* Riemann surface – a ("double") λ-plane with N removed points: in the roots of the polynomial (2.7). Then the formula (4.11) can be interpreted in the following way. The only non-perturbative effect in terms of this Riemann surface is blowing up its singularities by the simplest possible procedure – replacing the marked points at $\lambda = \phi_i$ by the "handles": $w + \frac{\Lambda^{2N}}{w} \sim \lambda - \phi_i$, and passing in this way from the λ-plane with marked points to a smooth Riemann surface (fig. 11).

A degenerate Riemann surface – "two copies" of the λ-plane with N marked points is depicted at the top of fig. 14. This degenerate limit, was already mentioned before, corresponds to weak coupling in $\mathcal{N} = 2$ supersymmetric gauge theory and, therefore, can be computed straightforwardly using one-loop perturbation theory. The opposite degenerate limit is much more interesting and corresponds to the degenerate Riemann surface in the bottom of fig. 14. This limit is stable when the extended supersymmetry is broken down to $\mathcal{N} = 1$ (the corresponding values of moduli of this degenerate curve are exactly in the minima of $\mathcal{N} = 1$ potential). It is this limit, when the periods (4.27) vanish (the **B**-cycles correspond to small circles on fig. 14 while the differential (4.28) does not have any singularities at corresponding points) and it means that the corresponding masses of magnetic monopoles also vanish in this limit. The effective $\mathcal{N} = 1$ superpotential acquires the form

$$\mathcal{W} = \tilde{Q} a^D(u) Q + \mu u \tag{4.31}$$

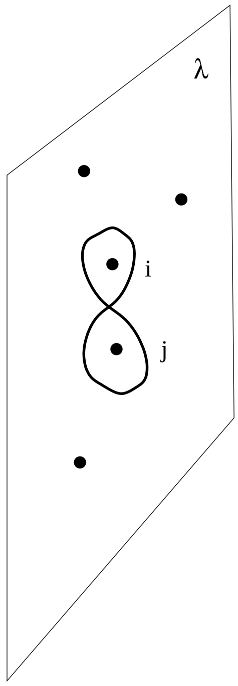

Figure 13: *"Figure-of-eight contour drawn around the points $\lambda = \phi_i$ and $\lambda = \phi_j$ in λ-plane, which is an analog of the A-cycle on the Seiberg-Witten curve.*

where $u = \langle \mathrm{Tr}\phi^2 \rangle$, Q and $\tilde{Q}$ are the vacuum values of the monopole supermultiplets and μ is the scale of violation of $\mathcal{N} = 2$ down to $\mathcal{N} = 1$. The function $a^D(u)$ is defined by the integral (4.27). It follows from here that in the minimum $\langle \tilde{Q}Q \rangle \sim \mu$, or the monopoles in $\mathcal{N} = 1$ theory condense at this leads to the (dual to well-known in superconductivity) effect when the electric field is "forced out", i.e. to (Abelian) confinement. Thus, the supersymmetric Seiberg-Witten theory becomes a nice "exactly solvable" laboratory for studying the properties of real QCD [43, 44].

4.6 Exact Nonperturbative Results and Integrable Systems

The fact that string theory possesses an extremely high symmetry allows one in practice for the first time to raise a question about the computation of the exact correlation functions in absolutely nontrivial theories, moreover not belonging formally to the class of quantum integrable models at least in canonical sense. The main idea of getting exact answers from symmetry considerations is based on deriving the relations, which correlation functions should obey. If the symmetry is high enough these relations may lead to the exact solution. It was in the framework of string theory (more strictly in the framework of its simplest models) that such program was completely carried out and it turned to be possible to get exact (in particular nonperturbative) information about the correlation functions.

First, some progress was achieved in the theories "without matter" or in the theories of two-dimensional gravity interacting with "minimal" ($c \leq 1$) matter (let us recall, that the central charge c counts the number of degrees of freedom). It turned out that such theories can be effectively described in terms of the matrix models of two-dimensional gravity [59], i.e. in terms of the *finite-dimensional* matrix integrals of the form

$$Z = \int DM \exp(-V(M)) \tag{4.32}$$

where $DM \propto \prod_{i,j} dM_{ij}$ denotes the simplest integration measure over the finite-dimensional matrices. The loop expansion or the expansion over topologies of the matrix graphs [68] of the integral (4.32) reproduces the (discretized version) of the loop expansion (3.11) of $c \leq 1$ string models. The double-scaling limit of the formula (4.32) [60] allows to identify $\mathcal{F} \propto \log Z$ directly with the full generating function of the string theory correlators

$$\langle \mathcal{O}_{i_1} \dots \mathcal{O}_{i_n} \rangle = \frac{\partial^n \mathcal{F}}{\partial T_{i_1} \dots \partial T_{i_n}} \tag{4.33}$$

and/or with the effective action. The information about the function $\mathcal{F}$ can be encoded in the set of nonlinear integrable equations.

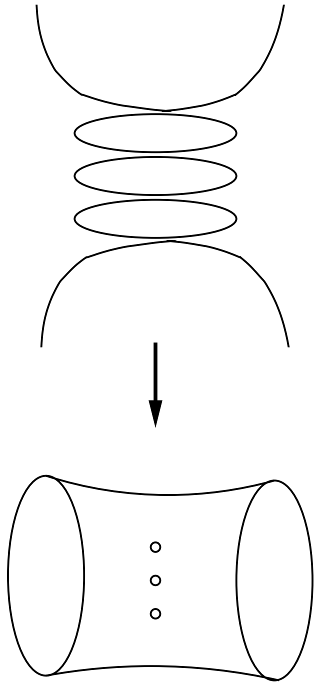

Figure 14: *Two degenerate limits of the smooth curve from fig. 11.*

The generating function depends on variables of two types. The first type of variables is the set of sources for physical operators

$$\mathcal{F}(g_{\rm str},\mathbf{T}) = \sum_{g=0}^{\infty} g_{\rm str}^{2g-2}\mathcal{F}_g(\mathbf{T}) = \\ = \sum_{g=0}^{\infty} g_{\rm str}^{2g-2}\int Dh_{ab}\ D\mathbf{X}\ \exp\left(-S_{CFT}(\mathbf{X},h_{ab}) + \sum T_k\mathcal{O}_k\right) \tag{4.34}$$

and the derivatives of (4.33) over these sources determine the correlation functions in the theory. Expression (4.34) does depend upon the choice of basis of the operators $\mathcal{O}_k$ or parameters T_k, and only in some fixed basis (not necessarily convenient from the point of view of the world-sheet theory) it can be elegantly described in terms of non-linear partial differential equations or unitarity-like relations for the correlators. In general, such relations are well-known in traditional quantum field theory (the Ward identities, the Schwinger-Dyson equations etc) but the situation in string theory is singled out by the fact that there equations can be written in the form of *closed* system of *integrable* equations completely fixing the generating function (4.34). As a function of parameters $\mathbf{T}$, the generating function (4.33), (4.34) can be defined only in the sense of formal series, whose coefficients are identified with the correlation functions, but the series itself has vanishing radius of convergency. This fact reflects the well-known properties of the perturbative expansions in string theory and quantum field theory and moreover it is consistent with the existing explicit formulas for the exact nonperturbative solutions. If they exist these formulas are usually known in the form of integral representations and may sometimes be written in terms of the matrix integrals (4.32). However, the particular terms of the series for (4.34), for example $\mathcal{F}(\mathbf{T}) \equiv \mathcal{F}_0(\mathbf{T})$ can be found and written in terms of well-defined functions.

Another set of parameters, which the partition functions or generating functions depend on, are the physical or space-time moduli of the theory. The space of these parameters is usually finite-dimensional, in the considered cases it is also often complex and may be interpreted as moduli space of complex manifolds. I repeat that complex curves or Riemann surfaces arising in this context have the "space-time origin" (say come out of the string compactification) and are not related to the world-sheets of string theory!

As a function of moduli the generating function is a normal (say, meromorphic) function of many complex variables and can often be computed more or less effectively. The moduli parameters can be interpreted as the low-energy values of the background fields (the Higgs scalar condensates, moduli of physical metric – the Kähler and complex structures etc) and as a function of moduli the function $\mathcal{F}$ has usually the sense of an effective action. The existing relation between the geometry of complex manifolds and integrable systems allows one to identify the functions $\mathcal{F}$ with solutions to nonlinear integrable equations.

In general the dependence upon the generating parameters and moduli is rather different [26] and both functions are independently interesting problems. For example, in the Seiberg-Witten theory now there exists a reasonable answer only to the first question [27], and it is very important that the Wilsonian effective action in the massless sector can be expressed (4.19) via a *function* of several complex variables. Thus, it is the knowledge of the function $\mathcal{F}$ as function of moduli and all its derivatives, say the expansion over the sources $\mathbf{T}$ which gives the most complete information about the theory.

[26] In topological two-dimensional gravity and in some topological string models (of the A_p-type), the dependences on the moduli t and the sources T almost coincide (the $(t+T)$-formula [104]).

[27] Something about dependence on generating parameters and an analog of the $t+T$-formula in Seiberg-Witten theory can be found in [121].

The effective theory can be formulated in terms of (a classical) integrable system. This formulation is universal in the sense that it does not depend on many properties of the "bare" theory. For example, it does not really depend even on the dimension of a bare theory: two-dimensional, four-dimensional, and even five-dimensional theories look absolutely similar from this point of view. Moreover, so obtained effective theories remind a lot the *topological* field theories. They possess many properties of two-dimensional topological field theories, though the "bare" theories are essentially multidimensional and, what is especially important, contain massless propagating particles.

Let us now list the main types of differential equations arising in nonperturbative string theory.

- **The "Virasoro" constraints** (more strictly – the Virasoro-like constraints) [101, 102, 103]. This is one more manifestation of the not yet clear duality between the world-sheet and space-0time structures. The "Virasoro constraints" arising in matrix models of two-dimensional gravity and topological theories have the general form

$$\mathcal{L}_n \exp(\mathcal{F}) = 0 \tag{4.35}$$

  where $\mathcal{L}_n$ are the differential operators in parameters $\{T_n\}$, forming the Virasoro algebra (3.17). Note that equations of this type already arise in some effective space-time formulations of string theory. In contrast to Virasoro generators of the world-sheet reparameterizations, the operators $\mathcal{L}_n$ in this context have a purely space-time interpretation.

  Solution to the constraints (4.35), can usually be expressed through the tau-functions of the hierarchies of integrable equations. Sometimes these tau-functions can be written in terms of the matrix integrals (about appearance of the Virasoro constraints in matrix models and the relation between the matrix models and integrable systems see, for example, [11]). For the generating functions, written in the form of matrix integrals (4.32), the Virasoro constraints follow from the loop equations or Ward identities $\langle \delta V \rangle = 0$ (the average is understood in the sense of partition function (4.32)), which are basically the simplest analogs of the Ward identities in gauge field theory.

- **The associativity equations** [100]. A nontrivial over-determined system of differential equations to the generating function $\mathcal{F}$, containing its third derivatives. Collecting the third derivatives into the matrices $||\mathsf{F}_i||_{jk} = \frac{\partial^3 \mathcal{F}}{\partial T_i \partial T_j \partial T_k}$, the associativity equations can be written in compact form [111]

$$\mathsf{F}_i \mathsf{F}_j^{-1} \mathsf{F}_k = \mathsf{F}_k \mathsf{F}_j^{-1} \mathsf{F}_i \qquad \forall \ i, j, k. \tag{4.36}$$

  Firstly the associativity equations were found in topological string models (where they follow from the crossing relations) but later it turned out that they show up in much more vast class of effective theories, for example in the Seiberg-Witten theory.

- **"Quasiclassical" integrable hierarchies**. These hierarchies usually arise on attempts to find exactly the tree-level or spherical contributions $\mathcal{F}(\mathbf{T}) \equiv \mathcal{F}_0(\mathbf{T})$. They are usually reduced to well-known dispersionless analogs of the hierarchies of Kadomtsev-Petviashvili or Toda lattice types. In a wider sense the quasiclassical hierarchies are applicable, say, to the description of the Seiberg-Witten theory: the prepotential $\mathcal{F}$ is logarithm of the tau-function of some nontrivial solution to quasiclassical hierarchy. The known solutions to quasiclassical hierarchies are related mostly to geometry of complex manifolds. One of the consequences of such a relation is the existence of so called "localization" or the residue formulas of the form

$$\frac{\partial^3 \mathcal{F}}{\partial T_i \, \partial T_j \, \partial T_k} = \text{res} \left(\frac{dH_i dH_j dH_k}{\Omega} \right) \tag{4.37}$$

  where dH_i are one-forms related to the variables T_i, and Ω is some "symplectic" two-form. One of the possible consequences of the residue formulas is the existence of associativity equations (4.36).

5 Strings and Duality between Gauge Theories and Gravity

5.1 Holography and Strings

One of the most interesting recent physical ideas in string theory is applying the "holographic principle" which allows to describe theory in full D-dimensional space-time (or in some part of this space-time) – in the so called bulk – in terms of the information encoded on its *boundary*. Such a possibility exists far from everywhere, since the bulk theory contains, in general, much more information than the theory on the boundary – the number of degrees of freedom of the bulk theory is much larger. Roughly speaking, the ratio of the number of degrees of freedom in the bulk of dimension D and on the boundary of co-dimension δ (usually $\delta = 1$) under the growth of characteristic size of the system L grows as $L^D/L^{D-\delta} = L^\delta$. Besides this fact, in traditional quantum field theory the field theory "inside" (say, the Green functions) is completely determined by the boundary theory only in quadratic or free case.

In contrast to quantum field theory, string theory necessarily contains gravity, in which the relation between the bulk and boundary theories seems to be completely different. One of the manifestations of this fact is the well-known *linear* connection between the entropy of the black hole and the area of the horizon, demonstrating that the number of degrees of freedom in gravity is proportional not to the volume, as one would expect from quantum field theory. Another side of the same phenomenon is known as the 't Hooft holographic principle [105]. According to this principle due to deviation of rays in gravitational field any point from the bulk can be "independently" projected to the boundary (see fig. 15).

String theory unifies "matter" (open strings) and gravity (closed). Moreover, as was already discussed in sect. 4.4, there are natural vacua in string theory where matter is localized on some hypersurfaces in space-time,

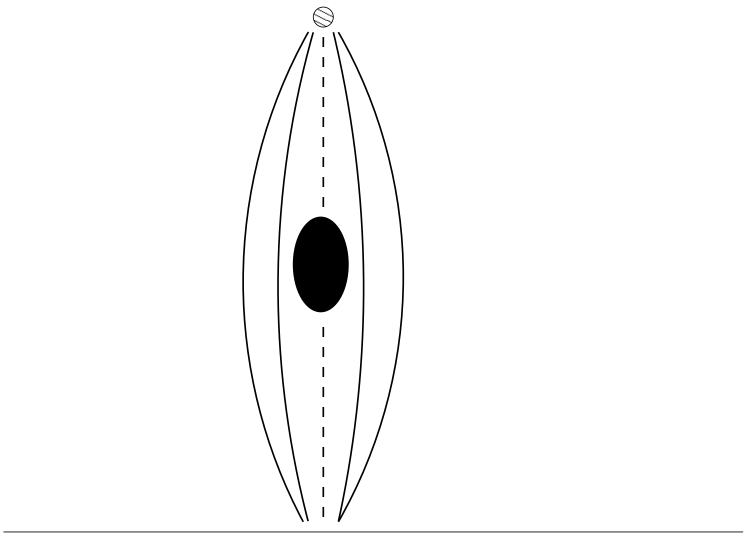

Figure 15: *Holographic 't Hooft principle. A point which cannot be naively projected to the boundary due to the presence of some material "screen", is nevertheless projected due to deviation of rays by the gravitational field, induced by this "screen".*

while gravitons or closed strings can propagate everywhere in bulk. A necessary production of closed strings in the theory of open strings (see fig. 5) leads to the possibility establishing some *holographic* (in the above sense) analogy between the theory of matter or open strings on a D-brane (on the 'boundary") and the theory of closed strings or gravity in the bulk.

In other words, the same effects can be formulated both in terms of open strings or the Yang-Mills theory as well as in the language of the closed string theory or gravity. In this chapter we will try to discuss some consequences of this duality, in the modern parlance usually called "AdS/CFT-correspondence", since the most well-known example of this phenomenon is the duality between $\mathcal{N} = 4$ supersymmetric conformal field theory of the Yang-Mills fields (conformal field theory – "CFT") and gravity in five-dimensional anti-de-Sitter space ("AdS") [122], see sect. 5.4 below. The most physically interesting effect which can hopefully be better understood in the framework of such correspondence is the parallel between two very important phenomena in modern theoretical physics proposed by Polyakov [117] – the confinement of quarks in non-Abelian gauge theories and the confinement of matter beyond the horizon of the black hole.

Another interesting aspect of this picture is adding to the physical picture of the world so-called "extra dimensions". In contrast to already traditional Kaluza-Klein ideas [62] (see, also, [42]) about additional *small* dimensions, responsible for the internal symmetries in the theory, in the new proposed physical picture the extra dimensions should not necessarily be small (and can in general be even non compact). The problems of the theories with extra dimensions (although not in the context of string theory) were considered recently in [38].

5.2 Duality of Open and Closed Strings

As we already discussed in sect. 3.1, string theory is the only reasonable candidate for the role of the unifying theory of the vector fields and gravity since it naturally unifies the carriers of these interactions as excitations of open and closed strings. One of the consequences of this relation is the possible interpretation of closed strings as bound states in the theory of open strings (see fig. 5). Another rather natural conclusion comes out if one considers the one-loop diagram in the open string theory corresponding to the world-sheet with topology of a cylinder (see fig. 16). Looking at the same diagram from the perspective of closed string theory, it is clear that it corresponds just to a tree-level propagator (cf. with fig. 6). Thus, it says that the one-loop (i.e. quantum) effects in the open string theory may have a dual formulation in terms of tree-level (i.e. classical) gravity – the massless part of the closed string spectrum.

This purely string duality can in principle be realized as a duality between the gauge theories and gravity and this leads to the already mentioned parallels between the confinement of quarks inside hadrons and keeping matter beyond the horizon of black holes. This idea has become very popular due to the more or less explicit example of the "holographic" duality between the $\mathcal{N} = 4$ supersymmetric gauge theory and geometry $AdS_5 \times \mathbf{S}^5$, or the direct product of five-dimensional Lobachevsky space or anti-de-Sitter space and five-dimensional sphere,

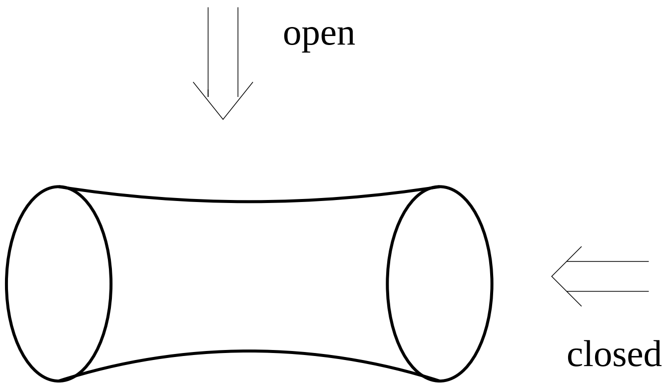

Figure 16: *One-loop diagram in the theory of open strings is equivalent to a tree diagram in the theory of closed strings.*

see sect. 5.4. Such duality is often called "holographic" since from the point of view of nonperturbative string theory one may consider it as a consequence of the holographic principle or, in more simple terms, of the fact that bulk gravity can be described in terms of some effective theory on the boundary of its volume. In more detail, the hypothetical scenario of such duality is based on the following assumptions:

- Matter, described in terms of gauge fields and their superpartners, or, generally, by open strings is confined to certain hypersurfaces in multidimensional (e.g. $D = 10$ or $D = 11$) space-time, since open strings are allowed to have their ends *only* on these Dirichlet or D-branes [28], see fig. 9.
- In contrast to matter, gravity corresponding to the massless excitations of *closed* strings, is allowed to propagate everywhere in the bulk of ten-dimensional space-time, i.e. is indeed (at least) ten-dimensional theory, as any consistent *quantum* gravity should be.
- The matter branes (D-branes) themselves induce a gravitational field, which, at the level of the classical ($\alpha' \to 0$) approximation could be considered just as a solution of the *bulk* equations of motion with the boundary terms arising from effective theories on branes. Hence, on the one hand one may look at the boundary terms induced by matter as at the (localized) sources for gravitational field, on the other hand deeper correspondence implies that gravitational boundary action may play the role of a generating function for the correlators in matter theory on brane.
- In most nowadays popular concrete models, the bulk geometry is "reducible" i.e. has form of a direct product like $AdS_5 \times S^5$, where the compact S^5 part is kept to be "fixed" while the real physics takes place within the other part, so that four co-ordinates $\{x_\mu\}$ play the role of "visible" space-time, while the rest, the fifth co-ordinate y (which the background metric nontrivially depends on), serves as a *scale* of observable space-time [29]. In other words the metric can be written in the distinguished in string theory form of the "Friedman universe"
$$ds^2 = dy^2 + a(y)(dx_\mu)^2 \tag{5.1}$$
- The scale factor of matter theory or the position of the matter brane in the auxiliary (fifth) dimension can be found as solution to the five-dimensional equations of motion (on the "gravitational side"), or by the renormalization group equations (on the "matter" or gauge theory side). Since equations of motion are differential equations of the second order (while conventional renormalization group contains only the first order equations in scale parameter), the relation between them is rather nontrivial. An interesting existing proposal is that of [126] – to use the Hamiltonian formalism [33] in five-dimensional gravity theory. Going along this way one should come to a direct description of the effective boundary action in terms of a tau-function of some integrable system, see sect. 4.6.

Most of these ideas about the relations between the gauge theories and theory of gravity arose [117] as a direct generalization of the well-studied correspondence between zero-dimensional (or one-dimensional) gauge theories – the so called matrix models (4.32) (or matrix quantum mechanics) and theory of two-dimensional gravity or $D \leq 1$ string models [59, 60, 97, 98, 99].

[28] At least in the context of type II string theory.

[29] In this context the five-dimensional geometry plays the role of the five-dimensional gravitational "bulk", restricted by "boundary" branes of codimension $\delta = 1$.

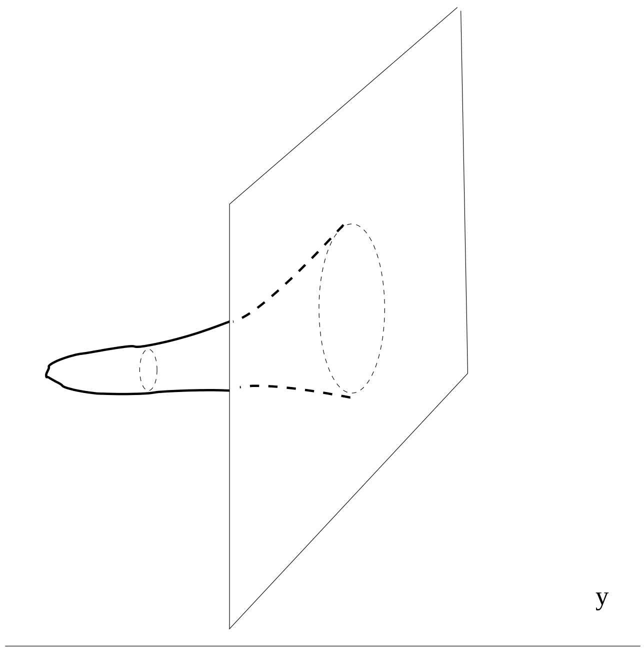

Figure 17: *Gravitational analog of confinement – the metric similar to the metric of a "black hole", which is almost flat far from the "horizon" and near the horizon turns into narrow throat with strong gravitational fields, confining "quarks".*

5.3 Confinement and Black Holes

One of the oldest problems in string theory, moreover being in a sense its main origin is the description of one-dimensional extended objects in the theory of strong interactions. Multiple attempts to formulate string theory adequate for the description of the Wilson loops in gauge theories and QCD has led to the idea [117] that such a theory should be necessarily noncritical in the sense that the effective tension must depend on auxiliary string co-ordinates playing the role of the scale factor and at some point this tension should vanish or become infinite. All that means that the string action in such model has the general form

$$\int_{\Sigma} \left(\partial\varphi\bar{\partial}\varphi + a(\varphi)\partial\mathbf{X}\bar{\partial}\mathbf{X} + \ldots\right) \tag{5.2}$$

in order to be able to coincide with the theory of gauge fields at critical point. The main problem then is to identify the action (5.2) with some exactly solvable two-dimensional conformal field theory with the necessary spectrum and other properties. In its main features the "gravitational picture" of confinement is depicted in fig. 17. The action (5.2) in gravitational approximation corresponds to the "Friedman metric" (5.1), the co-ordinates $\{x_\mu\}$ are the zero modes of "two-dimensional fields" $\{X_\mu(\sigma,\tau)\}$, while the co-ordinate y is the zero mode of the "two-dimensional field" $\varphi(\sigma,\tau)$. The function $a(y)$ qualitatively behaves in the following way: on one side of the y-axis it grows and the space-time becomes the macroscopic five-dimensional space. On the other side of the y-axis, contrarily $a(y) \to 0$, and one gets a "throat" with a strong gravitational field confining the matter.

The essential part of this picture is the "nonstandard" nature of gravity, compared to "ordinary" gravity of the observable (macroscopic) space-time. First, the effects of this "hadronic" gravity [117] should become essential not at the Planck scale but already at the scale of strong interaction of the order of 10^3 MeV. Second, metric (5.1) is not observable at least in the sense that the co-ordinate y is not a real co-ordinate or co-ordinate of "visible" space-time, but rather plays the role of a scale in the theory. Moreover, it is necessary to point out that the gravitational description is applicable only in the situation when string corrections are suppressed. It happens, for example, in the planar limit $N \to \infty$ [68], which corresponds to the tree-level Feynman diagrams of spherical topology or the spherical (i.e. tree-level) limit in dual closed string theory.

Thus, the existing examples of duality between gauge theory and gravity are implied to be correct at least in the phase where $N \gg 1$ and $g_{YM}^2 N > 1$. The first requirement is the well-known large N limit [68] and this means that in gauge theory only the planar diagrams survive, or that the loop corrections of the *closed* strings are suppressed. In contrast to this transparent limit of large N (which literally means $N \to \infty$ for the properly normalized quantities), while the second constraint onto the coupling constant is absolutely nontrivial. In order to compare it with the boundary action in the theory of gravity one should first sum up the contribution of all loops of the gauge theory or theory of open strings. Hence, the theory of gravity should predict the *nonperturbative* results in gauge theories which are not analytic in coupling constant. It is especially necessary to stress this circumstance in order to avoid mixing between the nontrivial string duality, relating the classical bulk theory with the boundary theory at strong coupling and rather trivial "continuation" of the (free) Green functions from the boundary. Such "continuation" is well-defined for conformal theory at the boundary and metric of constant negative curvature in the bulk.

5.4 AdS/CFT Correspondence

The most well-known example of duality between the gauge theory and gravity is the so called AdS/CFT correspondence – the correspondence of gravity in anti-de-Sitter space and the conformal field theory, or more exactly the $\mathcal{N} = 4$ supersymmetric Yang-Mills theory which is the four-dimensional (do not mix with two-dimensional) conformal field theory with the vanishing beta-function (2.4), (2.5) (at least in the perturbation theory). Such gauge theory can be represented directly by the picture at fig. 9, i.e. for the $SU(N)$ gauge theory – by a "stack" of N (completely coinciding!) D3-branes. The dual gravitational picture can be constructed as a solution to supergravity equations with corresponding boundary conditions. Such solution is well-known (see, for example, references in [122]), and its metric has the form

$$ds^2 = U^{-1/2}\left(dx_\mu\right)^2 + U^{1/2}\left(dr^2 + r^2 d\Omega^2\right) \tag{5.3}$$

while the source of this metric is the Ramond-Ramond 4-form

$$C_{\mu\nu\lambda\rho} = \epsilon_{\mu\nu\lambda\rho}\left(\frac{1}{U} - 1\right) \tag{5.4}$$

with the D3-branes being charged with respect to this form. In formulas (5.3) and (5.4) the function $U = U(r)$ depends only on the distance$r$ from the "stack" of branes

$$U(r) = 1 + \frac{g^2 N \alpha'^2}{r^4} \tag{5.5}$$

where N is the number of D-branes and g is the coupling of $\mathcal{N} = 4$ gauge theory. Metric (5.3) is a metric of manifold consisting of the five-dimensional sphere (the second term in the r.h.s. of (5.3)) and some five-dimensional manifold with the metric similar to (5.1), where the role of distinguished co-ordinate y is played by the distance r to D-branes. Since

$$U(r) \underset{r\to 0}{\sim} \frac{g^2 N \alpha'^2}{r^4} \tag{5.6}$$

in the vicinity of the horizon $r \to 0$, the first term in the r.h.s. of (5.3) turns into the anti-de-Sitter metric

$$ds^2 = \alpha'\sqrt{g^2 N}\left(\frac{dr^2}{r^2} + a(r)dx_\mu^2 + d\Omega^2\right) \tag{5.7}$$

where

$$a(r) = \frac{\alpha'}{g^2 N}\left(\frac{r}{\alpha'}\right)^2 \tag{5.8}$$

From (5.7) it follows that the squared radius of the five-dimensional sphere $R^2_{\text{sphere}} = \alpha'\sqrt{g^2 N}$ is equal to the so called 't Hooft constant in units of α'. As is was mentioned above, the string corrections are suppressed as $N \to \infty$, besides this requirement, metric (5.7) is close to the exact solution at large 't Hooft coupling, i.e. when $g^2 N \gg 1$.

This example is in fact the only *explicit* example of a relation between the gauge theory and gravity, which allows in particular to study the correlation functions and anomalous dimensions of composite operators [123]. Unfortunately this example cannot really be "deformed" into more sensible physical theories, i.e. all the construction is rather rigid. Some attempts of the dual gravitational description of the gauge theories with less supersymmetry were made in [127], though by now without any striking success.

From the more general point of view the AdS/CFT correspondence in the framework of string theory can be divided into two, generally speaking, different parts

$$\log \int DA_\mu \exp\left(-S_{YM}[A_\mu;\phi_0] + \sum \int_{d^4x} \phi_i O_i^{YM}(F_{\mu\nu})\right) =$$
$$= \sum \int D\varphi DX \exp\left(-\int_\Sigma \left(G_{MN}\partial X^M \bar{\partial} X^N + R^{(2)}\Phi(X) + \sum \phi_i(X)V_i(X)\right)\right) = \tag{5.9}$$
$$= \int_{dx} \sqrt{G}e^{-2\Phi}\left(R(G) + V(\phi_i) + \frac{1}{2}(\nabla\Phi)^2 + \frac{1}{2}(\nabla\phi_i)^2 + \ldots\right)$$

which are "labeled" by two different equality signs in the formula (5.9). This formula deserves further explanations which are now in order:

- The l.h.s. contains the logarithm of the generating function of the (supersymmetric, omitted for simplicity) Yang-Mills *matrix* field theory, which is considered in the sense of 't Hooft $\frac{1}{N}$-expansion, reproducing the perturbative expansion in string theory with both holes (open string loops) and handles (closed string loops, see fig. 8). One adds in this part the sum of the *gauge-invariant* operators $O_i^{YM}(F_{\mu\nu})$ [123] to the Yang-Mills action, depending on the (covariant derivatives of the) Yang-Mills field-strength with the *external* sources $\phi_i(x)$.
- The middle part is literally the string theory generating functional. As it should be in the first-quantized theory, there is the sum over topologies and number of "holes" (the Yang-Mills expansion we noted above). The integration is performed over all embeddings $X^M = (X^\mu, \varphi)$ of a two-dimensional world-sheet parameterized by (σ_1, σ_2) into the bulk space-time. By definition, the world-sheets may have holes only "attached" to the boundary in space-time, i.e. the Dirichlet boundary conditions have to be imposed on φ. The gauge invariant operators coupled to ϕ_i are now represented by the *closed-string* background fields $\phi_i(X)$, interacting with the string over the whole world-sheet surface.
- The requirement of two-dimensional conformal invariance (see sect. 3.4) is equivalent to the condition that external background fields $\phi_i(X)$ (including the specially singled out background metric $G_{MN}(X)$ and dilaton $\Phi(X)$) should be on a mass-shell, i.e. satisfy the equations of motion. This is an important point, because the equations of motion should be "supplemented" by boundary conditions, which are not explicitly mentioned in (5.9); nevertheless one should remember them and *add* to the "middle" part of (5.9) that the boundary conditions are imposed at $\varphi|_0 = y = y_*$ and the couplings in the Yang-Mills part (the l.h.s.) are exactly the *boundary* values of the string couplings $\phi_i(x) = \phi_i(X|_0, \varphi|_0 = y_*)$ [30].
- The equality between the middle part and the r.h.s. requires even more additional detailed explanations. The r.h.s. contains what is called the string theory effective action (see sect. 3.6; in particular eq. (3.26)). Literally as is written in (5.9) it looks like an ordinary low-energy effective action in quantum field theory. However, things are not so simple since one should remember that the middle part of the equality and, thus, the r.h.s. is defined *only* on mass shell. In fact the last part of formula (5.9) contains a *non-local* expression, arising if one substitutes into the action solutions to the equations of motion as functionals of the boundary conditions! Thus, despite it seems that formula (5.9) reformulates the quantum problem of computation of the generating function (taking into account all loop contributions) as some classical problem, the last one – the classical problem of finding the effective action as a functional of the boundary conditions – is not in fact simpler. An exception is the case of dilaton field with vanishing potential, where the comparison between the gauge theory and gravity was indeed performed in [123].

5.5 Life on a brane

Interpretation of the scale factor as an auxiliary co-ordinate of *space-time* allows one to consider the problems of confinement in the theory of elementary particles and the problems of gravity and cosmology on equal footing. In analogy to the previous section in the theory of gravity already at the level of simplest *classical* consideration it is easy to demonstrate [129] that

- It is easy to get a vanishing *effective* cosmological constant of the four-dimensional matter theory;

[30] We are now discussing this correspondence at a relatively "rough" level, forgetting more delicate questions, like the relation of the basis of gauge-invariant operators in the Yang-Mills theory and the basis of the corresponding vertex operator in string theory. This is a nontrivial issue, since there is no way to adjust these basises *a priori* in the first and second part of equality in the formula (5.9). This can be seen already for the simplest example of the AdS/CFT correspondence – the matrix model (4.32) and the dual theory of *two-dimensional* gravity.

- It is also easy to get a massless four-dimensional graviton, non propagating to the bulk at least in the linear approximation.

These two statements arise without any additional information from solving the Einstein equations of motion for bulk gravity with certain boundary conditions, induced by brane sources.

The most general classical action in this approach includes only two terms (the rest of contributions to the action are marked by dots)

$$\int_{d^5x} \sqrt{G_5}\left(\frac{R_5}{2\gamma_{\mathrm{N}}^{(5)}} + \Lambda_5\right) + \int_{d^4x} \sqrt{G_4}\Lambda_4 + \dots \tag{5.10}$$

where, according to the accepted rules, we consider only the nontrivial five-dimensional part of D-dimensional theory and write down two terms corresponding to the bulk five-dimensional contribution (with metric $G_{MN}^{(5)} \equiv G_{MN}$ and its curvature $R_5 = R_5(G)$; $\gamma_{\mathrm{N}}^{(5)}$ is the five-dimensional Newton constant) and the boundary four-dimensional contribution (where G_4 denotes the determinant of metric on the brane world volume, induced by the five-dimensional metric with the determinant $G_5 \equiv G$). The terms, omitted in (5.10), are generally non-local or contain higher derivatives; they however should necessarily be taken into account in an exact string formulation of the problem.

It is remarkable that the action (5.10) written in the simplest approximation, does not really depend on *any* details of the model. In the simplest case, the second term can be chosen as a δ-function along the fifth co-ordinate $x_5 = y$ and the "potentials" Λ_5 and Λ_4 can be considered as constants – the five-dimensional bulk cosmological constant and "bare" four-dimensional cosmological constants or tension of the correspondent brane. Nobody forbids, however, considering them as nontrivial functions of co-ordinates, being, say, the values of the matter (scalar) fields potentials – then the simplest picture is easily generalized to the case of several thin branes or a thick brane. The analysis in any case does not differ from the simplest examples of localization [128, 129], when the second term represents the only thin brane sitting at $y = 0$ with no other sources, or, better to say, the contribution of *all* other sources is encoded in the non-vanishing five-dimensional cosmological constant $\Lambda_5 = const < 0$ giving rise to the anti-de-Sitter AdS_5 geometry far outside the brane.

The appropriate solutions to the equations of motion, following from (5.10)

$$\frac{1}{\gamma_{\mathrm{N}}^{(5)}}\left(R_{MN}^{(5)} - \frac{1}{2}G_{MN}R_5\right) = \frac{1}{2}\Lambda_5 G_{MN} + T_{MN}^{(4)} \tag{5.11}$$

(in this section large indices run over five values $M, N = 1, \dots, 5$ while the small indices over the four values $\mu, \nu = 1, \dots, 4$) can be found in a very simple way, using the symmetries of the problem. Since $T_{MN}^{(4)} \sim \delta(y)t_{\mu\nu}^{(4)}(x)\delta_M^\mu\delta_N^\nu$ one can first solve eqs. (5.11) for $y \neq 0$, which naturally suggest the anzatz of a "Friedman universe" (5.1). Substitution of (5.1) into (5.11) gives

$$a''(y) + \frac{\Lambda_5\gamma_{\mathrm{N}}^{(5)}}{3}\, a(y) = 0, \qquad y \neq 0 \tag{5.12}$$

with the solution

$$\begin{gathered} a(y) = A\exp(ky) + B\exp(-ky) \\ \Lambda_5\gamma_{\mathrm{N}}^{(5)} = -3k^2 < 0 \end{gathered} \tag{5.13}$$

(the cosmological constant of five-dimensional space is negative). A natural choice would be $A = 0$ for $y > 0$ and $B = 0$ for $y < 0$, then we have an AdS horizon as $|y| \to \infty$. On the brane surface at $y = 0$ one has to "glue" two exponents with different signs, then $a(y) = e^{-k|y|}$, but this would bring us to an extra contribution into (5.11) at $y = 0$, i.e. proportional to $\delta(y)$. However, *tuning* $\Lambda_4\gamma_{\mathrm{N}}^{(5)} = 3k$ one exactly cancels this term by the contribution of the variation of the second term in (5.10) so that (5.11) also holds at $y = 0$. Thus, the solution is finally

$$ds^2 = \exp(-k|y|)(dx_\mu)^2 + dy^2 \tag{5.14}$$

so that the *effective* cosmological constant in four-dimensional theory

$$\Lambda_4^{\mathrm{eff}} = \Lambda_4 + \int dy\sqrt{G_5}\Lambda_5 = \Lambda_4 + \frac{\Lambda_5}{k} = 0 \tag{5.15}$$

vanishes. Thus, in this scenario the "observable" cosmological constant Λ_4^{eff} classically vanishes independently of any particular details of the model in a given class.

One of the *very* important immediate consequences we got in this context is that the boundary conditions (here – gluing on the brane) reduce exactly half of the bulk modes existing in the theory. In a more general context this condition could be different if speaking about its exact form, but one may always express in (5.13) B as a function of A or vice versa.

Next question to study is the spectrum of small fluctuations of the (linearized) action (5.10) in the vicinity of the background (5.14). It is easy to see that for the perturbation $g_{\mu\nu} = a(y)\eta_{\mu\nu} + h_{\mu\nu}(x,y) = a(y)\eta_{\mu\nu} + \psi^{(p)}_{\mu\nu}(y)e^{ipx}$ one gets an equation

$$\left(-\partial_y^2 + p_\mu^2 \exp(k|y|) - 2k\delta(y) + k^2\right)\psi^{(p)}(y) = 0 \tag{5.16}$$

rather similar to the Schrödinger equation in a δ-function well with a coefficient $-2k$. From elementary quantum mechanics it is well-known that there always exists a *single* level, localized to this well (here at $y = 0$) with the energy $E = -k^2$. This immediately gives rise to $p_\mu^2 = 0$ in (5.16), or to the four-dimensional *massless* graviton which is forbidden to propagate into the fifth direction (to the bulk) by the exponential wave function $\psi^{(p^2=0)} \sim e^{-k|y|}$.

This is, in fact, a generic phenomenon – for *any* metric of the form (5.1) with $a(y) = e^{-\alpha(y)}$ with suitable

$$a(y) \underset{|y|\to\infty}{\rightarrow} 0$$

there exists a solution to (5.11) with *non* constant bulk "potential" $\Lambda_5(y)$ and $\Lambda_4(y)$, corresponding in general to some thick brane, satisfying [31]

$$\begin{gathered}
\Lambda_5(y) = -3\alpha'(y)^2, \qquad \Lambda_4(y) = \frac{3}{2}\alpha''(y) \\
\Lambda_5(y) + \Lambda_4(y) = 3\left(-\alpha'(y)^2 + \frac{\alpha''(y)}{2}\right) \\
\int dy(\Lambda_5 + \Lambda_4)\exp(-2\alpha(y)) = \frac{3}{2}\int dy\frac{d}{dy}\left(\alpha'\exp(-2\alpha(y))\right) = \\
= -\frac{3}{4}\int dy\frac{d^2}{dy^2}\exp(-2\alpha(y)) = 0
\end{gathered} \tag{5.17}$$

Of course, the "gravity description" presented above answers almost all simple questions but cannot pretend to be complete [32]. The massive modes $\psi(y)$ can be expressed in terms of the Bessel functions and their contribution to the deviation from the Newton law in a four-dimensional world seems to be consistent with *any* one-loop contribution to the graviton propagator $\langle h_{\mu\nu}(x)h_{\alpha\beta}(0)\rangle$ which should be of the form $\int d^4q e^{iqx}k^4\log\frac{q^2}{\mu^2}$ giving rise to $\frac{1}{r^3}$ correction to the potential of four-dimensional gravity.

Now, let us remember that gravity arises only as an effective description of string theory and in the string theory picture the previous formulas can be understood in the following way. Consider the generating functional of string theory in the background (5.1), (5.2)

$$\int D\varphi DX \exp\left(-\int_\Sigma a(\varphi)\partial X_\mu\bar\partial X_\mu + \partial\varphi\bar\partial\varphi + \mathcal{R}^{(2)}\Phi(\varphi) + \dots\right) \tag{5.18}$$

so that the zero modes of $X_\mu(\sigma)|_0 = x_\mu$ play the role of four-dimensional co-ordinates in (5.1) while the zero mode of the Liouville field $\varphi(\sigma)|_0 = y$ is the extra bulk co-ordinate. The action (5.18) should be consistent in the sense of string theory, in particular after the integration over co-ordinates X_μ, the arising correction ГбвўЁо

$$\begin{gathered}
\int DX\exp\left(-\int_\Sigma a(\varphi)\partial X_\mu\bar\partial X_\mu\right) = \det\left(\bar\partial a(\varphi)\partial\right)^{-D/2} = \\
= \exp\left(-\int_\Sigma \partial\alpha\bar\partial\alpha + \mathcal{R}^{(2)}\alpha + \dots\right)
\end{gathered} \tag{5.19}$$

[31] Notice that the expression $\mathcal{T}(y) = \Lambda_5(y) + \Lambda_4(y)$ has exactly the form of the Miura stress-energy tensor, widely appearing in two-dimensional conformal theory, in particular in the procedure of bosonization or in the Liouville theory. Such "Virasoro" properties of the conformal mode of the space-time metric may serve as a possible origin for the target-space Virasoro symmetries (4.35), often appearing when describing the effective string theory actions in terms of integrable systems.

[32] For example, within pure gravity theory it is not clear why the classical vanishing of cosmological constant is not violated by quantum effects, say, by contribution of graviton tadpoles etc. This is just one more manifestation of the main concept of this review: the only way to "quantize" gravity is to consider it as low-energy limit of string theory.

should not break the conformal invariance (independence of the macroscopic theory of the choice of the world-sheet co-ordinates). In the last formula, which is a particular case of a general anomaly formula from [91], $\alpha = \alpha(\varphi) = \log a(\varphi)$, and the anomaly contributions depending only on metric are marked by dots. We see, that, identifying the Liouville or dilaton field with the fifth co-ordinate, eq. (5.19) gives rise to a reparameterization in the fifth dimension $\varphi \to \varphi + \alpha(\varphi)$ and $\Phi(\varphi) \to \Phi(\varphi) + \alpha(\varphi)$. For the particular background (5.14) one gets just a trivial *renormalization* of the string action for the Liouville component. In particular, this means that the background (5.14) is *stable* against string corrections. The integration over X_μ-coordinates is effectively equivalent to the study of nontrivial dependence only upon fifth coordinate in the bulk theory, taking four-dimensional branes as effective boundary sources and this is quite similar to what we have considered above in the classical gravity approximation. Moreover, the solution $\alpha = \alpha(\varphi) = \log a(\varphi)$ is the only one naively consistent with the requirement of world-sheet conformal invariance.

6 Some New Directions in String Theory

Finally in this review let us say a few words about the directions which have begun development only in recent years. We will discuss only few such directions and let us note immediately that the understanding of most of the problems considered in this chapter deserves to be better.

6.1 M(atrix) Theory

M(atrix) theory [118] is one of the most interesting (though not very successful) attempts to construct an alternative to strings formalism in M-theory. For the role of such formalism some particular *matrix* quantum mechanics is proposed. This origins already in its name and special attention to the first letter can be considered as a rather transparent hint that this letter should be identified with "M" in M-theory and the rest of the word "matrix" can be omitted.

As a building blocks m(atrix) theory uses the $N \times N$ matrices X_i, $i = 1, \ldots, 9$, whose diagonal elements can be interpreted as the transverse co-ordinates of the D0-branes (their number is equal to N) in the light-cone co-ordinates in the eleven-dimensional compactified M-theory. The Lagrangian of such a theory can be written in the form

$$\int_{dt} \mathrm{Tr} \left(\frac{1}{2R} \dot{X}_i^2 + M_{\rm pl}^6 R \sum_{i<j} [X_i, X_j]^2 + \ldots \right) \tag{6.1}$$

where the dots correspond to omitted fermionic terms. Eq. (6.1) explicitly contains the eleven-dimensional Planck mass $M_{\rm pl}$ (cf. with formulas (4.1) and (4.2)), together with the radius of the compact dimension R, which in the formalism of matrix theory somewhat artificially corresponds to the light-cone co-ordinate X_-. Hence, nine transverse co-ordinates and two light-cone co-ordinates – time and compactified X_-, corresponding to the trace over matrices in (6.1), together form the eleven-dimensional target space of M-theory.

The quantum mechanical action (6.1) can be interpreted in the following way. If $N = 1$, action (6.1) corresponds to the Hamiltonian $H \sim P^2$ and the ground state is degenerate with respect to all auxiliary (absent explicitly in (6.1)) Grassmann variables θ_α. Simple counting of all states shows (see, for example [28]), that their total number is $2^8 = 256$, so that half of them are bosonic: $44 = \frac{9(9+1)}{2} - 1$ gravitons and 84 of antisymmetric tensor field, and half of them are fermionic.

Thus, the "vacuum" of m(atrix) theory corresponds to the supergraviton, or, better to say, the supergraviton multiplet of eleven-dimensional supergravity [72], in which the only bosonic fields are metric and three-form. It is also claimed that nontrivial solutions to the equations of motion in m(atrix) theory can be identified with a membrane, fivebrane etc. For example, in the "quasiclassical" $N \to \infty$ action (6.1) can be rewritten, replacing the commutator with the Poisson bracket in auxiliary variables (σ_1, σ_2)

$$\int_{dt} \int_{d^2\sigma} \left(\frac{1}{2R} \dot{X}(\sigma_1, \sigma_2)_i^2 + M_{\rm pl}^6 R \sum_{i<j} \{X(\sigma_1, \sigma_2)_i, X(\sigma_1, \sigma_2)_j\}_{PB}^2 + \ldots \right) \tag{6.2}$$

This action can be identified with the action of membrane in the light-cone gauge.

The hope for a matrix formalism in nonperturbative string theory has not been justified in the sense that the new formalism appeared to be not very effective in solving essential problems. Nevertheless, it can be already considered as a "relative success" that at least some properties of string theory and eleven-dimensional

M-theory can be extracted from this at first glance totally absurd concept. To finish this section let us note, that some problems of the m(atrix) formalism were discussed in [12].

6.2 Non-commutative Field Theories

Non-commuting co-ordinates. The fact that the co-ordinates of the "stack" of D-branes from the point of view of effective field theory become eigenvalues of the matrix of scalar field in the adjoint representation is sometimes interpreted as appearance of *non-commuting coordinates*. In the first-quantized formalism one may consider this as a relatively simple and formal representation for the effective theories in terms of D0-branes, D-strings etc, studying the corresponding matrix quantum mechanics or two-dimensional non-Abelian gauge theory.

Non-vanishing background B-field. Another manifestation of non-commutativity shows up (see [130]) if we consider string theory in the nontrivial background B-field (3.4), for example

$$\begin{gathered} B_{\mu\nu} = B\epsilon_{\mu\nu} \\ B = \mathrm{const} \end{gathered} \tag{6.3}$$

This case can be clearly understood by analogy with the well-known example of a charged particle in a constant magnetic field. Indeed, the interaction, say, with the constant B-field (6.3), is performed over the whole surface of the world sheet

$$\int_\Sigma B_{\mu\nu} dX^\mu \wedge dX^\nu = \int_{\partial\Sigma} B_{\mu\nu} X^\mu dX^\nu \tag{6.4}$$

and by the Stocks formula it can be rewritten as a boundary term, equivalent to the interaction of a string with the vector-potential $A_\mu(X) = B_{\mu\nu}X^\nu$, corresponding to the constant magnetic field. If the value of the B-field is large enough the contribution of the term (6.4) to the two-dimensional correlator of the fields $X(t) = X|_{\partial\Sigma}$ dominates

$$\langle X_\mu(t) X_\nu(t')\rangle \propto \epsilon_{\mu\nu}\mathrm{sign}(t-t') \tag{6.5}$$

and in the field-theory limit this corresponds to non-commuting coordinates

$$[X_\mu, X_\nu] = \zeta\epsilon_{\mu\nu} \tag{6.6}$$

where $\zeta \sim \frac{1}{B}$. This reasoning is in fact a rather rough illustration of the well-known effect when the role of non-commutative variables is played by the centers of (small) circles – the trajectories of particles in a magnetic field.

The corresponding effective field theory can be described by a Lagrangian, where all the products are replaced with the so called Moyal products

$$f(x) * g(x) = \exp\left(\epsilon_{\mu\nu}\frac{\partial}{\partial x_\mu}\frac{\partial}{\partial y_\nu}\right) f(x)g(y)\Big|_{x=y} = f(x)g(y) + \{f(x), g(x)\} + \mathcal{O}(\partial^2) \tag{6.7}$$

where f Ë g are any two functions (local functionals) of "ordinary" fields $\phi(x)$, and

$$\{f(x), g(x)\} = \epsilon_{\mu\nu}\frac{\partial f}{\partial x_\mu}\frac{\partial g}{\partial x_\nu} \tag{6.8}$$

is the Poisson bracket, corresponding to the "quasiclassical" limit of the commutator (6.6). The Lagrangians where the fields are multiplied by the law (6.7), obviously contain infinitely many derivatives [33]. Examples of non-commutative field theories usually include the theories of scalar fields

$$S = \int_{dx}\left(\frac{1}{2}\partial_\mu\phi * \partial_\mu\phi + V(\phi)\right) = \int_{dx}\left(\frac{1}{2}\partial_\mu\phi\partial_\mu\phi + V(\phi)\right) \tag{6.9}$$

where $*$-multiplication (6.7) is essential only in the interaction terms, and the gauge theories

$$\begin{gathered} F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + A_\mu * A_\nu - A_\nu * A_\mu \\ S = \frac{1}{g^2}\int_{dx} F_{\mu\nu} * F_{\mu\nu} \end{gathered} \tag{6.10}$$

[33] Despite this, their ultraviolet properties are not better than the corresponding properties of ordinary, i.e. commutative quantum field theories.

which are rather natural generalization of the Yang-Mills theories. Notice, that in contrast to commutative case already the Abelian variant of (6.10) is a nontrivial interacting theory. Practically without any changes (just considering $A_\mu(x)$ as matrix-valued functions of non-commuting variables and adding the trace over matrix indices) formula (6.10) defines also the noncommutative Yang-Mills theories.

The most interesting by now applications of the non-commutative field theories are their classical solutions.

Solitons and instantons in non-commutative theories. In contrast to common scalar field theories where the existence of localized classical solutions is forbidden by scaling arguments in almost all dimensions (starting with $D \geq 2$), such solutions can arise in non-commutative field theories where the scaling is much less trivial due to an extra dimensional parameter (ζ in the formula (6.6)) [131]. The simplest is the two-dimensional case. After the scale transformation of co-ordinates $X \to \sqrt{\zeta}X$ in the action (6.9), one gets for the two-dimensional (or static three-dimensional) case

$$E = \int_{d^2x} \left(\frac{1}{2}(\partial\phi)^2 + \zeta V(\phi)\right) \tag{6.11}$$

and as $\zeta \to \infty$ the solution and its energy is completely determined by potential terms. The stationarity equation is reduced in such case, for example, for the potential $V(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{3}\phi^3$, to

$$m^2\phi + \lambda\phi * \phi = 0 \tag{6.12}$$

With normal multiplication, the solutions to (6.12) would be "maps into the set of points" $\phi(x) = 0$ and $\phi(x) = -\frac{m^2}{\lambda}$, however non-commutativity "washes away" these points in the space of fields. Indeed, formally a solution to (6.12) can be written as $\phi = -\frac{m^2}{\lambda}\hat{P}$, where $\hat{P}$ is the projector, i.e., generally, any operator with the property $\hat{P}^2 = \hat{P}$. In two-dimensional non-commutative space (isomorphic to the phase space of quantum mechanics with the only degree of freedom) projectors can easily be constructed in terms of, say, the Fock space operators. For example, one can take $\hat{P}_n \sim |n\rangle\langle n|$, where $|n\rangle$ – is the state of n-th energy level of harmonic oscillator. One can write correspondingly their representation in (non-commutative) x-space, the simplest solution will have a form of "bell" $\phi_0(x) = -\frac{2m^2}{\lambda} \exp(-(x_1^2 + x_2^2))$.

In the non-commutative gauge theories the main interest is caused by the instanton solutions [124]. In contrast to commutative theory, the nontrivial solutions to the self-duality equations arise already in the case of Abelian (noncommutative) group $U(1)$. From the physical point of view their main attraction is that they do not contain the singularities of the "zero-size" $1/x^4$ any longer (for example, in the expression for the field-strength at $\rho = 0$ in formula (2.12)), the parameter of non-commutativity turns the non-integrable singularity in four dimensions $1/x^4$ into the integrable expression $1/x^2(x^2 + \zeta)$. Construction of the solutions is almost the same as in the commutative case with the only distinction being replacement, as much as possible, of ordinary multiplication by the Moyal $*$-multiplication (6.7).

The detailed discussion of different aspects of the non-commutative theories can be found, for example, in the review [45].

6.3 Tachyon Potential

One of the main problems of many well-known string models is the presence of tachyons or states with negative squared masses. The tachyons lead, in particular, to infrared divergences in string amplitudes and since the infrared and ultraviolet regions are identified by two-dimensional geometry this problem "screens" the ultraviolet finiteness of string theory.

The interpretation of negative masses is absolutely clear in field theory (in particular, in the effective field theories for string models with tachyons) and it causes the *instability* of the corresponding vacuum. Indeed, drawing the effective potential with the requirement $m^2 = V''(\phi_0) < 0$, we immediately see (see fig. 18), that corresponding point (in the space of fields) is a local extremum but not a minimum, and under any perturbation the theory "runs" in the "true" vacuum at $\phi = \phi_*$.

Unfortunately, string theory by now does not have any self-consistent second-quantized formalism or string field theory [34] at least in the form, like the second-quantized approach exists in quantum field theory. Say, any

[34] The problems of constructing string field theory go beyond the scope of this review. Notice only that there exists a huge amount of literature, devoted to this problem, whose total volume can be easily comparable with amount of literature devoted to all other problems of string theory, taken altogether. The possibility of construction of string field theory is seriously restricted at least by absence of "universal variables" which allow to see *all* (and not a single!) string vacua, since the first-quantized string theory is described in terms of *different* variables – two dimensional conformal field theories – in the vicinity of different vacua. Another problem is that at least in the closed string theory the counting of states in the loops is different for a different number of loops (or for Riemann surfaces with different genera) and therefore, in contrast to quantum field theory, it is very hard (if possible!?) to write down a Lagrangian, taking into account this obstacle.

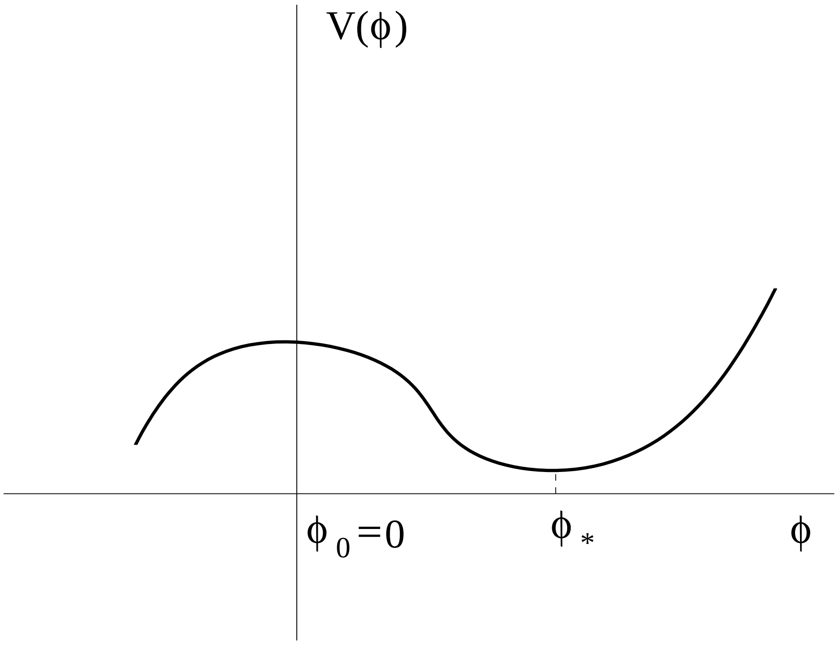

Figure 18: *Effective potential with minimum at $\phi = \phi_*$ and extremum at $\phi = \phi_0 = 0$. The value $\phi_0 = 0$ is an extremum point for the potential $V'(\phi_0) = 0$, but the second derivative is negative $V''(\phi_0) < 0$, what corresponds to the presence of the tachyon in the vicinity of this point.*

field theoretical Lagrangian with the potential depicted on fig. 18, allows one immediately to see both stable $\phi = \phi_*$, and non-stable $\phi = \phi_0$ vacua. This effect cannot be really seen in string theory since there is no formalism (yet?), which would allow one to consider the points $\phi = \phi_0$ and $\phi = \phi_*$ simultaneously.

In the bosonic string theory the existing formalism allows to compute amplitudes in the vicinity of vacuum of $\phi = \phi_0$ type, generally with two tachyons – from the open and closed spectrum. A.Sen [125] has proposed a nice D-brane interpretation, which allows partially to get rid of the tachyon of the open spectrum. It is based on the fact that the bosonic open string theory may be interpreted as D25-brane (the Dirichlet brane of dimension $p = 25$), whose world volume fills in the whole twenty-six-dimensional space-time. Equally the ten-dimensional superstring can be seen as a D9-brane. The standard way to get rid of the tachyon in ten-dimensional superstring – the GSO-projection [51], which was already discussed in sect. 3.5 – in fact corresponds to fig. 9 with parallel BPS D-branes. From some perspective this may even be considered as a definition of what is drawn on fig. 9.

Sen proposed to interpret the tachyon as a ground state of string, stretched between the Dirichlet and anti-Dirichlet branes, defining such a configuration as corresponding to the "opposite sign" in the GSO projection. It should be noted here that it corresponds only to the "non-diagonal" or "non-Abelian" tachyon of the open-string spectrum, since it corresponds to a string stretched between two different branes. Such a situation, in contrast to non-interacting parallel D-branes, is unstable. The Dirichlet and anti-Dirichlet branes tend towards each other and want to annihilate. From the energy conservation it follows that (see fig. 18)

$$V(\phi_0) - V(\phi_*) = 2T_D \tag{6.13}$$

where T_D – is the D-brane or anti-D-brane tension.

Moreover, since it is possible to stretch two strings between the D-brane and the anti-D-brane, different by orientation, the corresponding tachyon field becomes complex, and the potential from fig. 18 should be "complexified" by rotation around the vertical axis. Then it becomes similar to "bottom of a bottle" well-known in the framework of the Standard Model. The effective theory in such potential possesses "kink"-like solutions depending on some space-time co-ordinate x. For such solution one may take $\phi(x) \underset{x\to+\infty}{\to} |\phi_*|\exp(i\theta_1)$ and $\phi(x) \underset{x\to-\infty}{\to} |\phi_*|\exp(i\theta_2)$, with $\theta_1 \neq \theta_2$.

Hence, if the tachyon under discussion corresponds to the pair of Dp- and anti-Dp-branes, the arising kink is very similar to an extended object of a dimension less by unity, i.e. to a D$(p-1)$-brane. This kink is also unstable and it exists together with an "anti-kink" – a solution running along the co-ordinate x to the opposite direction. It is natural to interpret the anti-kink as an anti-D$(p-1)$-brane, and continue this procedure by induction. Such qualitative reasoning leads to the idea, that "falling down" along the tachyon potential depicted at fig. 18, from the point $\phi_0 = 0$ to the point $\phi = \phi_*$, and starting with a pair of Dp- and anti-Dp-branes, where $p = D - 1$ – is the dimension of our space (without time), we will find on our way many local extrema corresponding to the branes of smaller dimensions and finally will arrive at the "true" vacuum $\phi = \phi_*$, where the open string excitations are simply absent.

Unfortunately this sort of reasoning does not allow to compute the exact tachyon potential, even for restricted class of tachyonic fields. The only way to calculate such quantities is to use the effective actions which were discussed in sect. 3.6. Literally this method can be applied only in the vicinity of "false" vacuum $\phi_0 = 0$ of the tachyon potential, where corresponding two-dimensional conformal theory is a theory of free fields. However, there have been many attempts to "extrapolate" the results of such computations towards the direction of "real vacuum" $\phi = \phi_*$ (see, for example, [132]). Moreover, one can even find claims that the tachyon potential in tree-level approximation can be computed *exactly* [133], and equals to the rather simple expression (for the canonical kinetic term)

$$V(\tilde{\phi}) = -\frac{1}{2}\tilde{\phi}^2 \log \tilde{\phi} \tag{6.14}$$

with $\tilde{\phi} \sim \exp(-\phi)$. Despite the arguments in favor of this formula deserve to be more strict, qualitatively this means that in "true vacuum" $\tilde{\phi} = 0$ or $\phi \to \infty$ the mass of tachyon field becomes infinite, and it is consistent with the Sen hypothesis about the disappearance of all excitations of the open string spectrum.

7 Conclusion. String Theory or Field Theory?

In this review we have tried to discuss the main aspects of string theory in the form, as it exists at present. Certainly, as any physical theory detached from experiment it looks like it is "flying in the air" and the only excuses for such theory may come from new ideas, which have shown up inside string theory and, very slowly, affect the modern scientific paradigm of what is quantum field theory.

It becomes more and more evident that microworld physics cannot be simply reduced to an infinite set or "media" made of harmonic oscillators. Such theories arise only as the low-energy effective description of phenomena in the weak-coupling regime, which however finds lots of applications both in elementary particle and condensed matter physics. However, the main physical problems, which are not now understood, are contrarilyy related to the strong-coupling phase or strong filed regime, or exactly where the traditional quantum field theory or "theory of oscillator" does not have new successes. The very popular attempts thirty or even twenty years ago to develop "correct" or "general" formalism in quantum field theory, such that its computations can be "prolonged" towards the strong coupling look less and less promising. String theory in contrast implies (and originally implied) the existence of a principally new perspective on the problems of strong coupling.

Having appeared almost phenomenologically in the theory of strong interactions, the theory of one-dimensional extended objects gained huge popularity because, at variance with many other languages, it proposed a reformulation of many problems in terms of extremely simple two-dimensional conformal field theory, where the structure of computations is under the rigid control of infinite-dimensional symmetry and complex geometry, in particular by the language of complex analytic functions. Despite the observable world being multidimensional, the string scattering amplitudes are expressed through the correlation functions in two-dimensional conformal theories with well-defined operator product expansions etc. Moreover, the majority of target-space multidimensional symmetries are in this or that way related to the two-dimensional symmetries of the world-sheet theories.

In string theory the approach based on a dual description of the strong coupling effects was proposed and developed. Rather soon, this approach led to a certain hypothesis about nonperturbative results in supersymmetric gauge theories. These results are beyond the framework of traditional field-theoretical methods and allow one to get a deeper understanding of the problem of quark confinement.

String theory seems to be the only natural continuation of General Relativity to the region of strong fields and small distances. The (almost obvious) idea that there can be no quantum gravity in the framework of quantum field theory since these are two totally different theories is becoming more and more widespread. The appearance of time as a scale factor together with the distinguished role of solutions similar to the "Friedman universe" demonstrate deep internal relations between gravity and string theory.

Thus, the experience of development of string theory brought lots of rich new ideas into modern science. The only trouble is that string theory now does not possess not only a well-developed, but even any fixed formalism, allowing to perform computations of physical effects without applying of some "intuition". All these problems exist on the background of enforced development of connections with different spheres of mathematics and mathematical physics, and it allows to think that these problems have a temporary and mathematical, but not physical character. On the other hand, it is very nice to believe that it is a necessity to apply continuously physical intuition is called Theoretical Physics.

I am grateful to V.L.Ginzburg who proposed to collect several lectures and write a review to *Physics Uspekhi* and to S.Apenko, I.M.Dremin and L.B.Okun for reading the manuscript and useful comments. I am also grateful to E.Akhmedov, A.Barvinsky, I.Batalin, V.Ya.Fainberg, V.Fock, A.Gerasimov, A.Gorsky, Vl.Dotsenko, V.Kazakov, S.Kharchev, I.Krichever, A.Losev, V.Losyakov, Yu.Makeenko, R.Metsaev, A.Mironov, N.Nekrasov, A.Polyakov, A.Rosly, V.Rubakov, K.Saraikin, J.H.Schwarz, K.Selivanov, A.Tseytlin, I.Tyutin, A.Vainshtein, A.Yung, A.Zabrodin and, especially to B.Voronov and A.Morozov for lots of useful discussions of various questions considered in this text. I am also indebted to S.Apenko, H.Braden and E.Corrigan for useful advice concerning the English version of this review.

The work was partially supported by the RFBR grant No. 00-02-16477, the INTAS grant No. 00-00561 and the grant of support of scientific schools No. 00-15-96566. I would like to thank Laboratoire de Physique Théorique de l'École Normale Supérieure and Institut des Hautes Études Scientifiques, where this work was completed, for the warm hospitality

References

[1] A. Polyakov, *Gauge Fields and Strings*, Harwood Academic Publishers, 1987.

[2] M. Green, J. Schwarz and E. Witten, *Superstring Theory*, Cambridge University Press, 1987.

[3] J. Polchinski, *String Theory*, Cambridge University Press, 1998.

[4] A. Slavnov and L. Faddeev, *An Introduction to the Quantum Theory of Gauge Fields*, Moscow, Nauka, 1978.

[5] L. Okun, *Leptons and Quarks*, Nauka 1990

[6] I. Andreev, *Chromodynamics and hard processes at high energies*, Nauka 1981.

[7] P. West, *Introduction to Supersymmetry and Supergravity*, WS 1986.

[8] A. Marshakov, *Seiberg-Witten Theory and Integrable Systems*, World Scientific, 1999.

[9] V. Knizhnik, Physics Uspekhi **32** (1989) 945.

[10] A. Morozov, Physics Uspekhi, **162** (1992) 84-176.

[11] A. Marshakov, Int. J. Mod. Phys. A **8**, 3831 (1993), hep-th/9303101;
A. Morozov, Physics Uspekhi, **164** (1994) 3-62;
M. Adler and P. van Moerbeke, Commun. Pure Appl. Math. **50**, 241 (1997), hep-th/9706182.

[12] K. Zarembo and Yu. Makeenko, Physics Uspekhi, **168** N 1 (1998) 3.

[13] E. Akhmedov, Physics Uspekhi, **171** (2001) 1005-1024; hep-th/9911095.

[14] J. H. Schwarz, hep-ex/0008017; hep-th/0011078; hep-th/9812037; hep-th/9807135; hep-th/9711029.

[15] A. Sen, Nucl. Phys. Proc. Suppl. **94** (2001) 35-48, hep-lat/0011073; hep-th/9904207; hep-th/9802051.

[16] J. Polchinski, hep-th/9411028.

[17] J. Polchinski, hep-th/9611050.

[18] C. Vafa, hep-th/9810149; hep-th/9702201.

[19] P. Townsend, hep-th/ 9612121.

[20] J. H. Schwarz, hep-th/ 9607201.

[21] A. Losev, M-theory for pedestrians, preprint ITEP, unpublished.

[22] E. Verlinde and H. Verlinde, "Lectures On String Perturbation Theory," Published in Trieste School 1988: Superstrings:189 (QCD161:T7322:1988);
E. D'Hoker and D. Phong, Rev. Mod. Phys. **60** (1988) 917-1065
and references therein;
E. D'Hoker and D. Phong, hep-th/0110247, hep-th/0110283, hep-th/0111016, hep-th/0111040.

[23] C. Vafa, hep-th/9111017.

[24] E. Witten, hep-th/9112056.

[25] P. Aspinwall, hep-th/0001001.

[26] D. Morrison, math.ag/0007090.

[27] A. Marshakov, Theor. & Math. Phys. **112** (1997) 3-46, hep-th/9702083; Theor. & Math. Phys. **121** (1999) 179-243

[28] D. Bigatti and L. Susskind, hep-th/9712072.

[29] D. Bigatti and L. Susskind, hep-th/0002044.

[30] A. Giveon and D. Kutasov, Rev. Mod. Phys. 71 (1999) 983-1084, hep-th/9802067;
S. Kachru, hep-th/0009247;
P. Argyres and K. Narayan, JHEP **0103**, 047 (2001), hep-th/0101114.

[31] V. Ogievetsky and L. Mesincescu, Physics Uspekhi **117** (1975) 637.

[32] A. Vainshtein, V. Zakharov, V. Novikov and M. Shifman, Physics Uspekhi **136** (1982) 553-591; see also M. Shifman and A. Vainshtein, "Instantons versus supersymmetry: Fifteen years later," hep-th/9902018.

[33] L. Faddeev, Physics Uspekhi **25** 130 (1982)

[34] A. Morozov, Physics Uspekhi **150** (1986) 337-416.

[35] S. Gershtein, E. Kuznetsov and V. Ryabov, Physics Uspekhi **40** (1997) 773-806.

[36] V. Rubakov, Physics Uspekhi **169** (1999) 1299.

[37] G. 't Hooft, Physics Uspekhi **170** (2000) 1218-1224;
M. Veltman, Physics Uspekhi **170** (2000) 1225-1234.

[38] V. Rubakov, Physics Uspekhi, **171** (2001) 913-938; hep-ph/0104152.

[39] M. Vysotsky and R. Nevzorov, Physics Uspekhi, **171** N 9 (2001) 939-950.

[40] E. Likhtman, Physics Uspekhi, **171** (2001) 1025-1032.

[41] K. Wilson and J. Kogut, Phys. Rep. **12C** (1974) 75-199.

[42] J. Scherk, in *Recent Developments in Gravitation*, eds. M. Levy and S. Deser, Plenum, 1979.

[43] M. Strassler, Prog. Theor. Phys. Suppl. **131** (1998) 439; hep-th/9803009; *On confinement and duality*, Lecture notes for Trieste Spring School 2001.

[44] A. Yung, hep-th/0005088.

[45] M. Douglas and N. Nekrasov, hep-th/0106048.

[46] G. Veneziano, Nuovo Cim. **57A** (1968) 190.

[47] Y. Nambu, Proc. Int. Conf. on Symmetries and Quark Models, Wayne State Univ., Gordon and Beach 1970, 269-275;
T. Goto, Progr. Theor. Phys. **46** (1971) 1560.

[48] A. Neveu and J. Schwarz, Phys. Lett. **34B** (1971) 86.

[49] P. Ramond, Phys. Rev. **D3** (1971) 2415.

[50] J. Sбherk and J. Schwarz, Nucl. Phys. **B81** (1974) 118.

[51] F. Gliozzi, J. Scherk and D. Olive, Phys. Lett. **65B** (1976) 282.

[52] A. Polyakov, Phys. Lett. **B103** (1981) 207; Phys. Lett. **B103** 211.

[53] A. Belavin, A. Polyakov and A. Zamolodchikov, Nucl. Phys. **B241** (1984) 333.

[54] E. Witten, Commun. Math. Phys. **92** (1984) 455-472.

[55] V. Knizhnik and A. Zamolodchikov, Nucl. Phys. **B247** (1984) 83-103.

[56] M. Green and J. Schwarz, Phys. Lett. **149B** (1984) 117.

[57] E. Fradkin and A. Tseytlin, Nucl. Phys. **261** (1985) 1-27.

[58] A. Belavin and V. Knizhnik, JETP **91** (1986) 364-390.

[59] V. Kazakov, Phys. Lett. B **150** (1985) 282.
F. David, Nucl. Phys. **257** (1985) 45.

[60] V. Kazakov, Mod. Phys. Lett., **A4** (1989) 2125.

[61] J. Polchinski, Phys. Rev. Lett. **75** (1995) 4724, hep-th/ 9510017; Prog. Theor. Phys. Suppl. **123** (1996) 9, hep-th/ 9511157; Rev. Mod. Phys. **68** (1996) 1245, hep-th/ 9607050.

[62] T. Kaluza, Sitzungber. Preuss. Akad. Wiss. Berlin, Math.-Phys. KA, 1966 (1921);
O. Klein, Zs. Phys., **37** (1926) 895.

[63] Yu. Golfand and E. Likhtman, JETP Lett. **13** (1971) 452.

[64] J. Bekenstein, Lett. Nuovo Cim., **4** (1972) 737;
S. Hawking, Comm. Math. Phys. **43** (1972) 737.

[65] D. Gross and F. Wilczek, Phys. Rev. Lett. **30** (1973) 1343-1346;
H. Politzer, Phys. Rev. Lett. **30** (1973) 1346-1349.

[66] J. Wess and B. Zumino, Nucl. Phys. **B70** (1974) 39.

[67] G. 't Hooft, Nucl. Phys. **B79** (1974) 276;
A. Polyakov, JETP Lett. **20** (1974) 194.

[68] G. 't Hooft, Nucl. Phys. **B72** (1974) 461.

[69] A. Polyakov, Phys. Lett. **B59**(1975) 79; A. Belavin and A. Polyakov, JETP Lett. **22** (1975) 503-506.

[70] A. Belavin, A. Polyakov, A. Schwarz and Yu. Tyupkin, Phys. Lett. **B59** (1975) 85-87.

[71] E. Bogomolny, Sov. J. Nucl. Phys. **24** (1976) 449;
M. Prasad and C. Sommerfield, Phys. Rev. Lett. **35** (1975) 760.

[72] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. **76B** (1978) 409.

[73] N. Seiberg and E. Witten, Nucl. Phys. **B426** (1994) 19, hep-th/ 9407087; Nucl. Phys. **B431** (1994) 484, hep-th/ 9408099.

[74] J. Paton and H. Chan, Nucl. Phys. **B10** (1969) 516.

[75] J. -L. Gervais and B. Sakita, Nucl. Phys. **B34** (1971) 632.

[76] D. Volkov and V. Akulov, Phys. Lett. **46B** (1973) 109.

[77] L. Brink, P. Di Vecchia, P. Howe, S. Deser and B. Zumino, Phys. Lett. **B64** (1976) 435-438.

[78] S. Deser and B. Zumino, Phys. Lett. **B65** (1976) 369-373;
L. Brink, P. Di Vecchia and P. Howe, Phys. Lett. **B65** (1976) 471-474.

[79] M. Green and J. Schwarz, Nucl. Phys. **B181** (1981) 502; Phys. Lett. **109B** (1982) 444; Phys. Lett. **136B** (1984) 367.

[80] B. Feigin and D. Fuchs, Func. An. & Appl. **16** (1982) 47.

[81] Vl. Dotsenko and V. Fateev, Nucl. Phys. **B240** (1984) 312-348; **251** (1985) 691.

[82] D. Gross, J. Harvey, E. Martinec and R. Rohm, Phys. Rev. Lett. **54** (1985) 502.

[83] E. Fradkin and A. Tseytlin, Phys. Lett. **B163** (1985) 123.

[84] D. Friedan, E. Martinec and S. Shenker, Nucl. Phys. **B271** (1986) 93-150.

[85] N. Seiberg and E. Witten, Nucl. Phys. **B276** (1986) 272.

[86] M. Shifman and A. Vainshtein, Nucl. Phys. **B277** (1986) 456.

[87] E. Verlinde and H. Verlinde, Nucl. Phys. **B288** (1987) 357-379.

[88] E. Verlinde and H. Verlinde, Phys. Lett. **192B** (1987) 95;
J. Atick, G. Moore and A. Sen, Nucl. Phys. **B307** (1988) 221-273;
H. La and P. Nelson, Phys. Rev. Lett. **63** (1989) 24-27.

[89] J. Atick and E. Witten, Nucl. Phys. **B310** (1988) 291-334.

[90] V. Fainberg and A. Marshakov, Phys. Lett. **B211** (1988) 81-85.

[91] R. Kallosh and A. Morozov, JETP **94** (1988) 42-56.

[92] R. Dijkgraaf, E. Verlinde and H. Verlinde, Comm. Math. Phys. **115** (1988) 649;
P. Ginsparg, Nucl. Phys. **B295** [**FS21**] (1988) 153.

[93] H. B. Nielsen and D. Rohrlich, Nucl. Phys. **B299** (1988) 471.

[94] A. Alekseev, L. Faddeev and S. Shatashvili, Journ. Geom & Phys. **5** (1989) 391-406.

[95] A. Marshakov, Nucl. Phys. **B312** (1989) 178-196.

[96] A. Gerasimov, A. Marshakov, A. Morozov, M. Olshanetsky and S. Shatashvili, Int. J. Mod. Phys. **A5** (1990) 2495-2589.

[97] E. Brézin and V. Kazakov, Phys. Lett. **B236** (1990) 144.

[98] M. Douglas and S. Shenker, Nucl. Phys. **B335** (1990) 635.

[99] D. Gross and A. Migdal, Phys. Rev. Lett. **64** (1990) 127.

[100] E. Witten, Nucl. Phys. **B340** (1990) 281;
R. Dijkgraaf, H. Verlinde and E. Verlinde, Nucl. Phys. **B352** (1991) 59.

[101] M. Fukuma, H. Kawai and R. Nakayama, Int. J. Mod. Phys., **A6** (1991) 1385;
R. Dijkgraaf, E. Verlinde and H. Verlinde, Nucl. Phys. **B348** (1991) 435.

[102] A. Gerasimov, A. Marshakov, A. Mironov, A. Morozov and A. Orlov, Nucl. Phys. **B357** (1991) 565.

[103] E. Witten, in *New York 1991, Proceedings, Differential geometric methods in theoretical physics, vol. 1* 176-216;
A. Marshakov, A. Mironov and A. Morozov, Phys. Lett. **B274** (1992) 280;
S. Kharchev, A. Marshakov, A. Mironov, A. Morozov and A. Zabrodin, Nucl. Phys. **B380** (1992) 181, hep-th/ 9201013; see also [100].

[104] S. Kharchev, A. Marshakov, A. Mironov and A. Morozov, Mod. Phys. Lett. **A8** (1993) 1047, hep-th/9208046; see also [121].

[105] G. 't Hooft, gr-qc/9310026; see also [29].

[106] N. Seiberg, hep-th/ 9408013.

[107] A. Klemm, W. Lerche, S. Theisen and S. Yankielowicz, Phys. Lett. **344B** (1995) 169; hepth/9411048;
P. Argyres and A. Faraggi, Phys. Rev. Lett. **73** (1995) 3931, hepth/9411057.

[108] A. Gorsky, I. Krichever, A. Marshakov, A. Mironov and A. Morozov, Phys. Lett. **B355** (1995) 466; hep-th/ 9505035.

[109] P. Townsend, Phys. Lett. **B350** (1995) 184, hep-th/ 9501068.

[110] E. Witten, Nucl. Phys. **B443** (1995) 85, hep-th/ 9503124.

[111] A. Marshakov, A. Mironov and A. Morozov, Phys. Lett. **B389** (1996) 43, hep-th/9607109.

[112] S. Kachru, A. Klemm, W. Lerche, P. Mayr and C. Vafa, Nucl. Phys. **B459** (1996) 537-558, hep-th/9508155.

[113] E. Witten, Nucl. Phys. **B460** (1996) 335, hep-th/ 9510135.

[114] D. -E. Diaconescu, hep-th/ 9608163;
A. Hanany and E. Witten, hep-th/ 9611230.

[115] N. Seiberg and E. Witten, hep-th/ 9609219.

[116] E. Witten, hep-th/ 9703166.

[117] A. Polyakov, hep-th/9711002; hep-th/9809057; hep-th/0006132; hep-th/0110196.

[118] T. Banks, W. Fischler, S. Shenker and L. Susskind, Phys. Rev. **D55** (1997) 5112-5128; see also T. Banks, Nucl. Phys. Proc. Suppl. **67** (1998) 180, hep-th/9710231 and references therein.

[119] A. Marshakov, M. Martellini and A. Morozov, Phys. Lett. **B418** (1998) 294, hep-th/ 9706050;
A. Marshakov, in *New Developments in Quantum Field Theory*, Plenum Press 1999, NATO ASI Series B: Physics **366** (Eds. P. H. Damgaard and J. Jurkiewicz), 279; hep-th/ 9709001.

[120] B. Dubrovn, I. Krichever and S. Novikov, *Integrable systems - I*, in VINITI, *Dynamical systems - IV* (1985) 179;
N. Hitchin, Duke. Math. Journ. **54** (1987) 91;
A. Gorsky and N. Nekrasov, hep-th/9401021.

[121] A. Gorsky, A. Marshakov, A. Mironov and A. Morozov, Nucl. Phys. **B527** (1998) 690-716; hep-th/9802007.

[122] J. Maldacena, Adv. Theor. Math. Phys. **2** (1998) 231, hep-th/9711200.

[123] S. Gubser, I. Klebanov and A. Polyakov, Phys. Lett. B428 (1998) 105-114, hep-th/9802109.

[124] N. Nekrasov and A. Schwarz, Commun. Math. Phys. **198** (1998) 689-703, hep-th/9802068.

[125] A. Sen, hep-th/9805170; hep-th/9911116; hep-th/9902105.

[126] J. de Boer, E. Verlinde and H. Verlinde, JHEP **0008** (2000) 003, hep-th/9912012;
E. Verlinde and H. Verlinde, JHEP **0005** (2000) 034, hep-th/9912018;
E.Verlinde, Class. Quant. Grav. **17** (2000) 1277-1285, hep-th/9912058.

[127] J. Polchinski and M. Strassler, hep-th/0003136;
J. Polchinski, Int. J. Mod. Phys. A16 (2001) 707-718, hep-th/0011193.

[128] V. Rubakov and M. Shaposhnikov, Phys. Lett. **B125** (1983) 136.

[129] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (1999) 4690-4693, hep-th/9906064.

[130] N. Seiberg and E. Witten, JHEP **9909** (1999) 032, hep-th/9908142.

[131] R. Gopakumar, S. Minwalla and A. Strominger, JHEP **5** (2000) 20, hep-th/0003150.

[132] A. Tseytlin, Theor.& Math. Phys., **128** (2001) 540-560.

[133] A. Gerasimov and S. Shatashvili, hep-th/0009103.