The Large N Approximation to QCD

Carl Turner

Department for Applied Maths and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge, CB3 OWA, UK

E-mail: cpt39@cam.ac.uk, carl@suchideas.com

Abstract: This essay was submitted as a part of the *Part III Mathematics Examination* at the University of Cambridge in 2013. It is a review of the classic development of the so-called large N limit of QCD - that is, the theory of quarks and gluons with a large number of colours, or a large gauge group.

Note that it is fairly narrow in its remit - not touching the other applications of large N limits to more sophisticated theories, and not really exploring the modern fields of research spawned by the approach. We focus on understanding two main aspects: the historical evolution as led by 't Hooft, and a more modern, mathematical analysis by Gross and Taylor. We refer the reader to good places to read about some modern developments in 4D theory, including lattice theory and the omnipresent AdS/CFT correspondence, and point out how they develop the ideas we see here. However, this essay should be considered a nuts-and-bolts introduction to help get to grips with the large N formalism for the very first time.

Please do point out the mistakes and oversights which doubtless crept in (partly due to time constraints, partly due to restricting scope and length, and partly due to my stupidity)! I hope you enjoy reading it as much as I enjoyed learning enough to write it!

Contents

Typeset using JHEP styles

1 Introduction

It is a well known fact that the SU (3) gauge theory of *Quantum Chromodynamics* or *QCD* with quarks is a successful model of the strong nuclear interaction. However, many theories with gauge groups like SU (N) are subject to infra-red instabilities[1] which lead to the reign of non-perturbative phenomena at low energies. These, whilst necessary to explain the absence of free quarks (or more technically confinement of colour charge), mean that the perturbative quantum field theoretic techniques so heavily used in electroweak theory are not immediately applicable. This is because the coupling constant, whose behaviour as one *descends* from high energies is determined by renormalization group equations like (2.4), increases as one approaches terrestrial energy scales[1]. Thus it cannot be used as an expansion parameter.

The desire to find a new 'small' parameter is what led to Gerard 't Hooft's now famous 1974 paper[3] in which he described an approximation to an SU (N_0) gauge theory by considering the gauge theory SU (N) as $N \to \infty$. By rigging the coupling constant g such that $g^2 N$ is fixed, a resummation of the Feynman diagrams effectively gives an expansion in the parameter $1/N$. It is widely believed that fully interacting quantum field theories in four dimensions are not likely to be analytically tractable, so one hopes any such expansion reproducing *qualitatively* correct properties in some limit ($N \to \infty$ for us, as compared to the usual '$g = 0$' free case) will provide at the very least valuable insight into the theory away from this limit.

Since QCD is the theory of SU (3), and $1/3$ is not a very small number, one might be suspicious of the power of this theory. Yet, as we will discuss, the qualitative relevance of this way of viewing the theory seems almost indisputable. Further, some results are $\mathcal{O}\left(1/N^2\right)$ and so might correspond to a fairly low 10% error.

In section 2.3 and elsewhere, we will discuss several interesting hints of the behaviour of realistic QCD occurring in simpler models based on using this large N expansion to make calculations more tractable. In particular, despite the fact that we still generically cannot evaluate these sums of diagrams, even the structure of the expansion allows progress in understanding the physical states of QCD. Further, in 1+1 dimensional QCD, we note that the theory is essentially solvable.

However, perhaps a more interesting side-effect of this resummation is that diagrams of a particular order are characterized in a highly non-trivial way by their *topology*, rather than numbers of vertices; we find a series of diagrams strikingly like those from string theory. This seems key to a much investigated link between low-energy QCD and string theories, which we will develop from two perspectives (and refer to in a third).

Our programme is as follows:

§2 The Large N Approximation We describe the motivation for and nature of the particular large N approximation considered by 't Hooft. We also look at how some experimental facts are reflected by our framework.

§3 The 1+1 't Hooft Solution Next, we look at the application of this approximation to 1+1 dimensional QCD (in flat space). We explain the emergence of a discrete spectrum of meson states and the lack of a spectrum of free quark states in the presence of antiquarks. Then we look briefly at some other results arising from the formalism, including scattering amplitudes for these mesons. We then move on to attempt to gain some understanding of the link between QCD and string theories, first looking at a straightforward approach by Itzhak Bars to write a theory

[1] Lattice calculations suggest this is a reasonable deduction, even though the RG equation (2.4) is perturbative.[2]

involving strings which reproduces QCD calculations. We see that Bars's approach reproduces 't Hooft's meson spectrum and other aspects of his solution.

§4 The 1+1D Case in a Non-Trivial Background Pursuing our investigation of the relationship with string theories, we remove quarks but generalize to a non-trivial 1+1 dimensional spacetime, to see how the geometrical structure of a string theory emerges to all orders in N^{-1} in a more subtle case.

§5 Conclusions Finally, we review the material we have discussed, and assess its relevance to the study of QCD. In particular, we ask how well the large N approximation, the 1+1D cases studied and string theories in general correspond to real, four dimensional SU (3) gauge theory. We also look at contemporary attempts to expand the string analysis of the theory via gravitational duality.

The aim of the essay is to provide an introduction to the general framework, and then see how it opens up several wholly new types of calculation in strongly coupled gauge theories. Therefore, we provide an overview of the salient points along with the key mathematical parts of the arguments in each case, but focus on obtaining qualitative understanding from our calculations. This is explored in detail in the conclusion, especially section 5.3.

We draw especially upon the original papers of 't Hooft[3, 4], and papers by Callan[5] and Bars and Green[6–8], for the first two sections. Gross and Taylor's paper[9] is our main resource for the section on their theory.

The author would like to thank Professor Michael B. Green for helpful suggestions on the structure, level of detail and emphasis of the essay.

2 The Large N Approximation

The system we are approximating is QCD, which of course has an $\mathrm{SU}(3)$ gauge symmetry, so the obvious choice of theory to investigate is $\mathrm{SU}(N)$ gauge theory with an anti-Hermitian gauge vector field $A_\mu = -(A_\mu)^\dagger$ coupled to, say, three quarks q^a. However, we note that we have a relation for the generators of this gauge group of the form

$$\mathrm{SU}(N): \quad \left(T^A\right)^i{}_j \left(T^A\right)^k{}_l = \frac{1}{2}\delta^i_l \delta^k_j - \frac{1}{2N}\delta^i_j \delta^k_l \tag{2.1}$$

where the second term occurs to reflect the fact that the generators are traceless. By contrast, the larger group $\mathrm{U}(N)$ has the structurally simpler relation

$$\mathrm{U}(N): \quad \left(T^A\right)^i{}_j \left(T^A\right)^k{}_l = \frac{1}{2}\delta^i_l \delta^k_j$$

which differs only by a term of order $1/N$. (Of course, the normalization is arbitrary.) This motivates simplifying the Feynman rules slightly by considering this latter gauge group, and hoping that the one extra generator has negligible effects in this limit[2]. The propagator for the gluon fields is proportional to this term; looked at another way, we should add an extra $\mathrm{U}(1)$ ghost field precisely cancelling diagrams due to the $\mathrm{U}(1)$ field in the $\mathrm{U}(N)$ gauge group. In fact, it turns out (see Appendix A on page 34) that the effects of this extra gauge boson are suppressed by a factor of $1/N^2$, and so can we can neglect this correction and still hope for results of up to 10% accuracy, good for our expansion and even comparisons with experiment.

Remark. We will return to the case of $\mathrm{SU}(N)$ in section 4 where we investigate a remarkable relationship with a string theory; we will also see a different interpretation of using $\mathrm{U}(N)$ instead. Nonetheless, our current formalism can be straightforwardly extended to $\mathrm{SU}(N)$; see for instance [5].

2.1 The Lagrangian and Feynman Rules

We make the usual definition of the field strength tensor and covariant derivative

$$\begin{aligned} G_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu + g\left[A_\mu, A_\nu\right] \\ D_\mu &= \partial_\mu + g A_\mu \end{aligned}$$

where the quarks transform in the fundamental representation of the group, so that $\left(A_\mu q^a\right)_i \equiv \left(A_\mu\right)_i{}^j q^a_j$. Then the Lagrangian is

$$\mathcal{L} = \frac{1}{4}\mathrm{tr}\left(G_{\mu\nu}G^{\mu\nu}\right) - \bar{q}^a\left(\gamma^\mu D_\mu + m_{(a)}\right) q^a \tag{2.2}$$

where we allow the quarks to have different masses $m_{(a)}$. Note we are using a $(-+++)$ convention.

To obtain Feynman rules in the Feynman gauge, we introduce a Faddeev-Popov ghost ϕ in the adjoint representation and add the following terms to the Lagrangian:

$$\mathrm{tr}\left(\frac{1}{2}\partial_\mu A^\mu \cdot \partial_\nu A^\nu - \partial^\mu \phi^\star \cdot \left[\partial_\mu \phi + g\left[A_\mu, \phi\right]\right]\right)$$

[2] This extra 'photon-like' particle represents an extra ninth gluon field, arguably an importance difference. The reason for making this choice - also made by 't Hooft in the original paper [3] - is that introducing the condition that the gauge field is traceless forces us to include explicit dependence on N in the Feynman rules, something we would rather avoid for simplicity. See Appendix A on page 34.

In the interests of adhering to [3], we will take $x_\mu = (\mathbf{x}, x_d = it)$ in d dimensions and so the metric is simply $\delta_{\mu\nu}$.

The two sets of charges for the complex fields, A_μ (off the diagonal) and ϕ, mean we must introduce two distinct arrows into the Feynman diagrams tracking each set of indices. In Figure 2.1, we use a large arrow to show flow of ϕ-charge. For the colour indices, which we may consistently write up and down in some order as in $A_i{}^j{}_\mu = -A_j{}^i{}_\mu^\star$ so that upper indices are contracted with lower ones[3], we use a trick of 't Hooft to simplify the diagrams.

Instead of thinking about N^2 different gluons, we observe that a gluon field is specified by an ordered pair of colours (i, j), with the antihermiticity condition giving a real field for $i = j$ and complex fields, with conjugate schematically given by $-(j, i)$, otherwise. Thus we write fields in the adjoint representation with two arrows travelling in opposing directions, letting upper indices be incoming arrows, and lower indices outgoing. We fix the overall direction such that the gluon propagator in Figure 2.1 represents a right-moving $A_i{}^j{}_\mu$ for $i > j$ and a left-moving $A_j{}^i{}_\mu$ field if $i < j$. Then, since all indices are always contracted at vertices, we can forget about this ordering, and just join up the two sub-lines in such a double line with single quark propagators and other constituents of double-lines, ensuring the arrows flow continuously so that the matrix multiplications make sense. (The vertices in Figure 2.1 illustrate this.)

We follow the conventions for tracking is of 't Hooft for ease of comparison with his work; this means that in particular loop momenta correspond to integrals

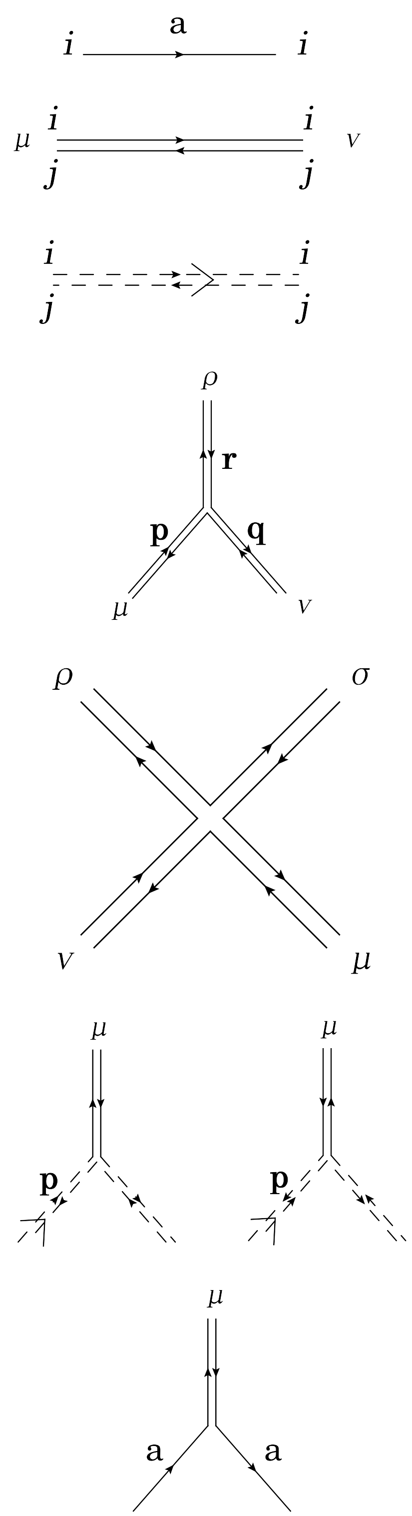
$$\int \frac{\mathrm{d}^d p}{(2\pi)^d\, i}$$

Remark. We will not draw on most of these rules in this essay, since the only computation we complete working directly with the theory is in a very special case. However, we reproduce them here from [3] for completeness.

This set of rules has the appealing feature that the rules do not depend explicitly on N, which is good if we want to find a theory where '$N = \infty$'. Nonetheless, amplitudes will contain multiplicative factors of N arising from loops transporting a gauge index i, since $\delta^i_i = N$, which will cause manifest divergences. This is physically obvious, since a finite-amplitude process which can happen in infinitely many ways (i.e. in any colour) will correspond to an infinite amplitude overall. (See section 2.3 for a different way of tracking Ns.)

To fix this divergence, we must modify the amplitudes to counterbalance these growing powers of N. The only logical free parameter we have is the coupling g, so we consider taking $g \propto N^{-k}$ for some k; by choosing k large enough we hope to find diagrams in a series with amplitudes of decreasing powers of N. So how should we choose the scaling?

[3]See for instance equation (2.1). This is possible because the Lagrangian consists of matrix multiplications and traces.



$$\frac{1}{m_{(a)} + i\gamma \cdot k - i\epsilon}, \text{ quark propagator}$$

$$-\frac{\delta_{\mu\nu}}{k^2 - i\epsilon}, \text{ gluon propagator}$$

$$\frac{1}{k^2 - i\epsilon}, \text{ ghost propagator}$$

$$ig\left(\delta_{\nu\rho}\left(r-q\right)_{\mu} + \delta_{\rho\mu}\left(p-r\right)_{\nu} + \delta_{\mu\nu}\left(q-p\right)_{\rho}\right)$$

$$g^2\left(2\delta_{\rho\mu}\delta_{\sigma\nu} - \delta_{\rho\sigma}\delta_{\mu\nu} - \delta_{\sigma\mu}\delta_{\rho\nu}\right)$$

$$igp_{\mu}, \qquad -igp_{\mu}$$

$$-g\gamma_{\mu}$$

Figure 2.1: The Feynman rules (omitting momentum conservation, loop integrals and the -1 for fermion loops). Colour indices are always transported trivially along coloured lines, and are omitted from the interactions. The large arrow on the ghost indicates the flow of ghost charge.

2.2 Simple N Counting

Firstly, note that if we assume that colour indices are not transported in or out of the diagram by currents at external vertices (all sources and observables are gauge-invariant, or roughly speaking, observers cannot see colour charge[4]) then the power-counting of N, g in a diagram is fairly simple. Consider an arbitrary connected diagram, with all index-carrying lines drawn on. We consider the propagators for A_μ and ϕ to be *single edges*, of double lines, for our purposes. Each of the V_3 3-point vertices in this diagram have a factor of the coupling g, whilst the V_4 4-point vertices all have a factor g^2.

Now, the power of N in the diagram is given by I, the number of *index loops*. We construct a surface by attaching a face for each index loop, making faces share an edge along double lines. However, now each quark edge is only attached to one side of a face. 't Hooft originally[3] proceeded by adding an extra patch to each *quark loop*, of which there are L. Applying Euler's formula for a surface without boundary would now lead to the result we want.

This is, however, a slightly inconvenient way of looking at the problem. The reason for this is that the surface we have constructed could be non-orientable; but note that the double lines of the gluon propagators define an orientation consistent amongst faces separated by gluon propagators, by insisting on an anticlockwise arrow along the border. Therefore, instead we define an orientable surface *with a boundary consisting of quark loops* by stitching together polygons defined by edges with gluon propagators.[5] Now the faces of this polyhedron is $F = I$, and we instead have $B = L$ boundaries. See Figure 2.2 for an example; we see diagrams have the scaling

$$g^{V_3+2V_4} N^I \propto N^{F-k[V_3+2V_4]}$$

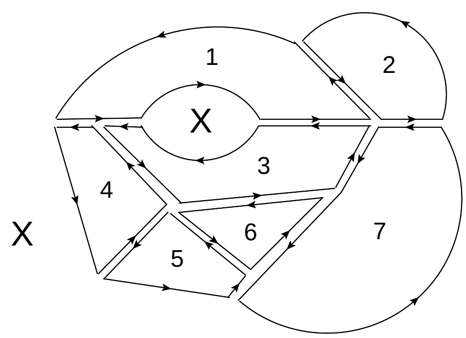


Figure 2.2: An example of N counting, displaying $F = I = 7$ numbered faces and $B = L = 2$ excluded regions, marked with Xs, on an $H = 0$ plane.
Here $V_3 = 10$, $V_4 = 2$, $I = 7$, so we have a scaling $g^{14}N^7 \equiv \left(g^2 N\right)^7 \times N^0$, where $7 \equiv \frac{1}{2} \cdot V_3 + 1 \cdot V_4$ and $0 \equiv 2 - B$.

Now an elementary calculation making use of Euler's formula for the Euler characteristic χ, that is $V - E + F = 2 - 2H - B \equiv \chi$ (where H is the genus of a surface with B boundaries containing a

[4] This has long had not simply experimental support, but also mathematical evidence in the theory suggesting that 'coloured states' have infinite energy[10], and even speculative attempts to show QCD may be thought of as an inherently colourless theory[11].

[5] This is the method followed by Coleman[12, p. 371] and Manohar[13], except that the latter defines rescaled fields which we only use in section 2.3; it has the advantage of agreeing with the string interpretation discussed later.

connected graph of $V = V_3 + V_4$ vertices, E edges and F faces) we find our answer as follows:

$$F = \chi + E - V, \qquad E = \frac{1}{2}(3V_3 + 4V_4)$$
$$F - k(V_3 + 2V_4) = \chi + \frac{1}{2}(V_3 + 2V_4)(1 - 2k)$$

Thus $k \geq \frac{1}{2}$ is sufficient to lead to a sensible-looking expansion. However, this expression for the N-dependence also suggests a preferred choice of k, namely $k = \frac{1}{2}$. We see that taking

$$g^2 N = \text{const.}, \qquad N \to \infty$$

gives a (connected) diagram of L quark loops and H handles (when thought of as defining a surface with L boundaries) the weight

$$\boxed{N^\chi \times \left(g^2 N\right)^{\frac{1}{2}V_3 + V_4} \qquad \propto \qquad N^{2-2H-L}} \tag{2.3}$$

in our expansion, so that we naturally categorize our diagrams by topological considerations, in a manner very reminiscent of string expansions for a coupling $1/N$.

Remark. Note that, as observed in [13], the $g^2 N = $ const. rigged limit can also be derived from consideration of the renormalization group equations for QCD[1]; one finds

$$\mu \frac{\mathrm{d}g}{\mathrm{d}\mu} = -\left(\frac{11}{3}N - \frac{2}{3}N_{\text{flavours}}\right)\frac{g^3}{16\pi^2} + \mathcal{O}\left(g^5\right) \tag{2.4}$$

so that taking $N \to \infty$ with g constant leads to an imbalanced equation. Since the two sides scale as g and Ng^3 to leading order, it is natural to take $g \sim g^3 N$ or $g^2 N \sim$ const. in order to obtain a theory running sensibly for large N. Concretely, writing $g_0 = gN^{1/2}$ as the constant coupling,

$$\mu g_0'(\mu) = -\left(\frac{11}{3} - \frac{2}{3} \cdot \frac{N_{\text{flavours}}}{N}\right) g_0^3 + \cdots$$

which suggests that effects with quark loops (which are where we get sums over flavours) are $1/N$ suppressed, in agreement with the N^{-L} part of the scaling shown in equation (2.3).

We will not explore other possible limits, including allowing $N_{\text{flavours}} \propto N$, and taking the quarks in different representations; the reader is referred to the literature for examples.[14, 15]

In the following, we are most interested in properties of quarks, so that $L \geq 1$; vacuum-to-vacuum gluon diagrams are irrelevant when using quark currents. In this case, the choice $\{L = 1, H = 0\}$ gives the leading diagrams. The $H = 0$ condition implies the diagram can be drawn on a sphere with $L = 1$ excluded regions. Remove the quark loop temporarily, project the graph onto the plane and reinstate the quark loop as a boundary for the whole graph; then

$$\boxed{\text{the leading diagrams with quarks are planar, with the quark loop around the exterior}}$$

which makes them look like simple open string diagrams in string theory, with ends at the quarks. An example is shown in Figure 2.3.

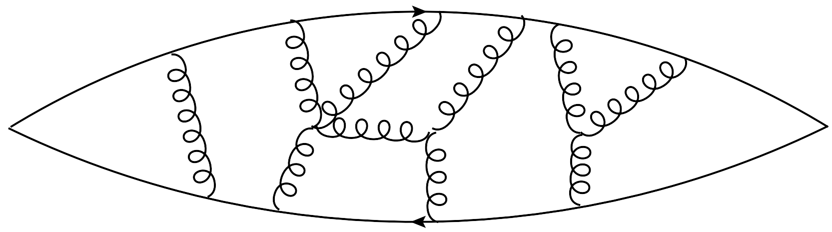


Figure 2.3: Traditional form (with double-line structure obscured) of a typical leading order quark loop diagram with $\{L = 1, H = 0\}$.

2.3 N-Counting for Correlation Functions

The above analysis is the most obvious way to obtain simplified 'N-counting rules' for a typical graph. Yet we wish to extend our above analysis from to vacuum expectation values of gauge-invariant operators, for which it is useful to use the following more elegant approach, advocated by Manohar[13]. We find it convenient to define rescaled fields

$$\hat{A} = A/\sqrt{N}, \qquad \hat{\psi} = \psi/\sqrt{N}$$

so that the Lagrangian before gauge fixing (2.2) becomes

$$\mathcal{L} = N\left[\frac{1}{4}\mathrm{tr}\left(\hat{G}_{\mu\nu}\hat{G}^{\mu\nu}\right) - \bar{\hat{q}}^a\left(\gamma^\mu \hat{D}_\mu + m_{(a)}\right)\hat{q}^a\right]$$

and all g terms are replaced with g_0 in $\hat{G}, \hat{D}$. Thus the N-counting rules become simply

$$\begin{aligned} \text{propagator} \quad &\longleftrightarrow \quad \frac{1}{N} \\ \text{vertex} \quad &\longleftrightarrow \quad N \\ \text{index loop} \quad &\longleftrightarrow \quad N \end{aligned}$$

and then connected vacuum diagrams, drawn on the surface constructed in the previous section, manifestly come with a N^χ factor - there are E propagators, V vertices and F index loops, and $\chi = V - E + F$.

The advantage of this new format, with a new set of fields, is that it generalizes easily to correlation functions. Suppose $\hat{O}_i$ are suitable operators written in terms of the new fields - namely, assume they are gauge invariant, and cannot be factorized into colour singlets like $\bar{\psi}_i\psi_i\bar{\psi}_j\psi_j$ for $\psi = \hat{q}^a$ (so that we do not violate the *connected* condition). Then we can preserve the N-counting rules by adding current terms $NJ_i\hat{O}_i$ to $\mathcal{L}$, since this is simply a new vertex with the correct N behaviour. The logarithm of the partition function, $W(J)$, generates connected diagrams with these currents,

$$\left\langle \hat{O}_1 \cdots \hat{O}_r \right\rangle_{\mathrm{C}} \propto \frac{1}{N}\frac{\delta}{\delta J_1}\cdots\frac{1}{N}\frac{\delta}{\delta J_r}W(J)$$

where the subscript C refers to the connectedness. Consequently,

$$\left\langle \hat{O}_1 \cdots \hat{O}_r \right\rangle_{\mathrm{C}} \propto N^{-r} \times N^{\chi_0}$$

where χ_0 is calculated for the leading order diagram possible for this process. For purely gluonic expectations, $\chi_0 = 2$ whilst for expectations involving quarks $\chi_0 \leq 1$.

Let $\hat{Q} \sim \bar{\psi}\psi$ be a quark bilinear in the rescaled $\hat{q}$ field, and $\hat{B}$ a purely gluonic operator in the rescaled $\hat{A}$ field. Then schematically

$$\left\langle \hat{Q}\hat{Q} \right\rangle_{\mathrm{C}} \sim N^{-1} \qquad \text{and} \qquad \left\langle \hat{B}\hat{B} \right\rangle_{\mathrm{C}} \sim 1$$

Hence to obtain canonically normalized propagators, operators like $\sqrt{N}\hat{Q}$ create *meson states* whilst those like $\hat{B}$ create *glueball states*.

This is enough to conclude the main results of interest. Firstly, s-meson scattering amplitudes go like

$$\left\langle \sqrt{N}\hat{Q}_1 \cdots \sqrt{N}\hat{Q}_s \right\rangle \sim N^{1+\frac{1}{2}s-s}$$

so that any scattering process ($s \geq 3$) is suppressed by powers of N (bounded by $1/\sqrt{N}$). Secondly, interactions between glueballs and mesons are bounded by

$$\left\langle \sqrt{N}\hat{Q}\hat{B} \right\rangle \sim N^{1-\frac{1}{2}-1} \sim 1/\sqrt{N}$$

Therefore we conclude that, if it is confining, large N QCD ought to be a *weakly interacting theory of mesons (and glueballs)* with a coupling $1/\sqrt{N}$. In particular, in the $N \to \infty$ limit we expect a *free* spectrum of mesons and glueballs.

One can go much further with this sort of analysis. We recommend [13] for an overview of the many applications to meson phenomenology. One can deduce some interesting information purely from N-behaviour; for instance

- chiral symmetry breaking can be established[16] and the relative importance of terms in chiral perturbation theory predicted fairly accurately;
- the absence of 'exotic states' such as a bound $q\bar{q}q\bar{q}$ can be partially explained by their absence at leading orders in $1/N$ - one can see this heuristically by noting that their propagator

  $$\left\langle \left(\sqrt{N}\hat{Q}_1 \sqrt{N}\hat{Q}_2 \right)^2 \right\rangle \sim 1/N$$

  or much more carefully by constructing this propagator for a $qq\bar{q}\bar{q}$ state and noting that it is a *free pair of mesons* to leading order[13]; and
- *Zweig's rule*, which states that certain meson interaction diagrams - where the mesons can be separated by only cutting gluon lines - are suppressed, has at least a partial explanation in this framework (though see [5, 17]), since this requires more quark loops than the leading order diagram.

However, we now move on to attempt to gain a more thorough, mathematical understanding of the theory, which requires some simplifying assumptions. The key case we consider is QCD in 1+1 dimensions.

3 The 1+1 't Hooft Solution

Field theories in 1+1 dimensional spacetime have been intensively studied due to the fact that many substantial simplifications occur in lower dimensions.[6] This often leads to exact solutions, as in the case of bosonization[18] of one-dimensional fermionic systems.

Gerard 't Hooft found[4] that the simplifications afforded by working in the *light-cone gauge* in 1+1 large N theory allowed the summation of the planar diagrams relevant to understanding quark and meson behaviour. The key idea is to use coordinates $x^{\pm} = \left(x^1 \pm x^0\right)/\sqrt{2}$ to obtain a metric

$$\mathrm{d}s^2 = 2\mathrm{d}x^+\mathrm{d}x^-$$

and then to use the gauge freedom to enter light-cone gauge with

$$A_- = A^+ = \frac{1}{\sqrt{2}}\left(A_0 - A_1\right) = 0$$

This results in

$$G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + g\left[A_\mu, A_\nu\right] \quad \longrightarrow \quad G_{+-} = -\partial_- A_+$$

with other components zero or related to this by antisymmetry. Clearly the gluon fields, now encoded entirely in $\left(A_+\right)_i{}^j$, are no longer self-interacting. Moreover, the 'kinetic' terms for the gluon fields are (using the metric and antisymmetry)

$$\frac{1}{4}\mathrm{tr}\left(G_{\mu\nu}G^{\mu\nu}\right) = \frac{1}{4}\mathrm{tr}\left(G_{+-}G^{+-} + G_{-+}G^{-+}\right) = -\frac{1}{2}\mathrm{tr}\left(\left[\partial_- A_+\right]^2\right)$$

Hence we can in fact take these fields to be *non-dynamical*![7] Just as in the case of electromagnetism, the non-dynamical field is expected to create some sort of Coulomb force between the particles sourcing the field, the quarks. Note that Coulomb forces in D spacetime dimensions have Green's functions of the form $|\mathbf{x} - \mathbf{y}|^{3-D}$ (giving rise to a familiar inverse-square law in $D = 4$). In fact, a model of Schwinger[18, 19] shows that an abelian theory (essentially QED) can demonstrate properties related to confinement in 1+1D, so it will perhaps not be a surprise to uncover confinement in the now effectively Abelian theory of QCD in two dimensions. However, it is instructive to understand how the results of our expansion result in a discrete meson spectrum at low energies. We will return to questions of the usefulness of the two-dimensional approach in section 5.3.

Note that in this gauge, there is no ghost.[4] Now expanding the covariant derivative term in the Lagrangian density gives

$$-\bar{q}^a\left(\gamma\cdot\partial + m_{(a)} + g\gamma_- A_+\right)q^a$$

so that the quark propagator and quark-gluon vertices, neglecting the colour index management, take the forms

$$\frac{1}{m + i\gamma\cdot k - i\epsilon} = \frac{m - i\gamma_- k_+ - i\gamma_+ k_-}{m^2 + k^2 - i\epsilon} \qquad \text{and} \qquad -g\gamma_-$$

[6] There are various reasons; most amount to the fact tensors have many fewer free parameters, so for example, spin has a different nature here, and there are also no transverse directions in light-cone gauge. This latter observation indicates that massless fields - which generically have $(D-2)$ polarizations in D dimensions - are very simple here. See section 5.3 for a discussion of the usefulness of the model.

[7] The idea is that one may take x^+ to be the 'time direction' so that no time derivatives act on A_+, and hence its equations of 'motion' take the form of constraints. This means A_+ can be solved for given the other fields. There is here the usual technical subtlety of light-cone gauge, namely that we assume the invertibility of the operator ∂_- - see [6] and below. We continue optimistically.

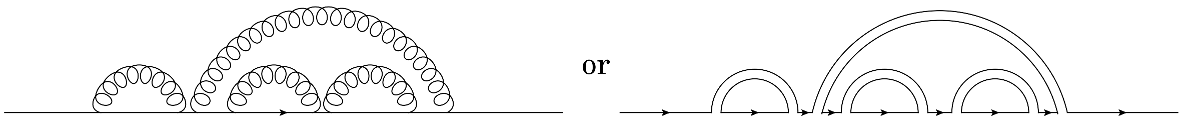


Figure 3.1: A typical 'rainbow' correction to the quark propagator.

respectively. However, taking the quark propagators to always terminate in these vertices, it is clear that we can take advantage of the fact that $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}$ to simplify these to

$$\frac{-i\gamma_+ k_-}{m^2 + k^2 - i\epsilon} \quad \text{and} \quad -g\gamma_- \qquad \equiv \qquad \frac{-ik_-}{m^2 + k^2 - i\epsilon} \quad \text{and} \quad -2g$$

respectively, using first $\gamma_-^2 = 0$ and then $\gamma_-\gamma_+\gamma_- = \{\gamma_-, \gamma_+\}\gamma_- - \gamma_+\gamma_-\gamma_- = 2\gamma_-$. (The assumption that quark propagators always end in these vertices means we can effectively cancel the γ_- in this last equation.)

Since there are no gluon-gluon interactions, the rules for this theory are much simpler than those developed for the theory above. Further, at leading order, the absence of fermion loops and planarity requirement mean that the geometry of diagrams for fermion propagators is essentially a series of straight fermion lines running through the diagram with gluons passing between them without crossing. Generically these take the form of so-called *ladder diagrams* - we either have loops on individual lines, or 'rungs' crossing between adjacent ones. See Figure 3.3 for an indication of how these rungs appear in summations.

Crucially for what follows, we note that the leading diagrams for the one quark loop case have (as noted prominently in section 2) only an *exterior* quark loop. This translates to all the loops on a single quark (which are self-energy corrections discussed in the next section, by the planarity of the diagram) being on the *same side* of the quark line, as seen in Figure 3.1, and as we will discuss in the next section slightly more generally.

Remark. The validity (and indeed singularity) of this axial gauge was a concern to many at the time; in [6] Bars and Green address these issues with a more rigorous calculation than we present here. They demonstrate the essential equivalence of their theory with 't Hooft's more naïve model, including Poincaré invariance and gauge invariance for colour singlets, so we stick to the simpler calculation.

3.1 Quark Propagators

As a special case, we begin by considering the self-energy diagrams for fermions responsible for the renormalization of the quark mass. These are formed solely by gluons looping away from the line and returning to it, retaining the planarity requirement. We emphasize the importance of the planarity constraint again here, as it allows the simplification noted at the end of the last section.

Fact. *The planar self-energy corrections to quark propagators in our vacuum diagrams due to gluon loops take the form of 'rainbow diagrams' like Figure 3.1, in which we can take all gluon loops to fall on the same side of the quark line.*

This is clear, since the quark line, which must be part of a quark loop, forms a boundary, and hence one of the two regions isolated by this quark loop on the constructed surface must be omitted. But a planar diagram must then have all gluon loops on the non-omitted patch (otherwise they must be stuck on as a 'handle').

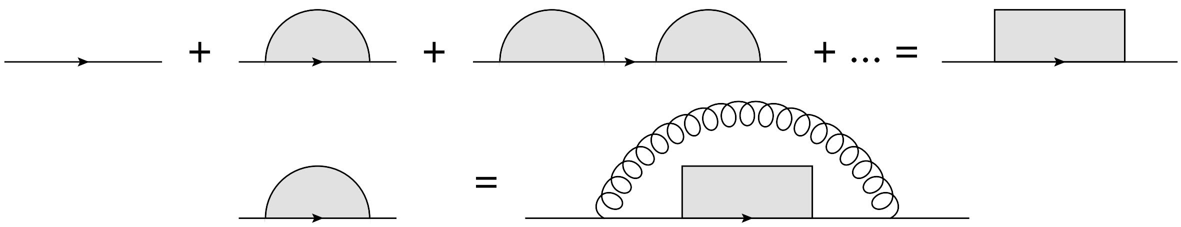


Figure 3.2: Deriving the bootstrap equation for the propagator. Here, the semicircle represents all 1PI contributions, whilst the box represents the full propagator.

Now the renormalization of the mass is expected to modify the above propagator via

$$\frac{-ik_-}{m^2+k^2-i\epsilon} \quad\longrightarrow\quad \frac{-ik_-}{m^2+k^2-k_-\Gamma(k)-i\epsilon}$$

where $i\Gamma(k)$ is the sum of 1PI diagrams[8] contributing to the self energy.

Remark. To see why it takes this form, note that the most general correction to the propagator is a chained sequence of these 1PI graphs as depicted in the first line of Figure 3.2 - so letting Π_0 be the undressed propagator we find a *dressed propagator*

$$\Pi = \Pi_0 + \Pi_0 i\Gamma\Pi_0 + \Pi_0 i\Gamma\Pi_0 i\Gamma\Pi_0 + \cdots = \frac{\Pi_0}{1-i\Gamma\Pi_0}$$

or

$$\frac{-ik_-}{m^2+k^2-i\epsilon-(-ik_-)(i\Gamma)} = \frac{-ik_-}{m^2+k^2-i\epsilon-k_-\Gamma}$$

as stated above.

Further, the form of the infinite sum of 1PI self-energy graphs can easily be self-consistently determined as follows: there must be a gluon loop; and the first gluon loop l_1 to leave the quark line must be the *last* to return, since otherwise planarity requires a gap between the point where l_1 meets the quark line and where the next gluon loop leaves, making the diagram one-particle reducible. But given this, any correction to the propagator between l_1 leaving and returning is possible.

Hence the 1PI graph contribution is of the form of one gluon loop containing a copy of the full dressed propagator, as shown in the second line of Figure 3.2. Hence, putting all of Figure 3.2 together, we bootstrap our way to an integral equation for Γ. Consider a quark propagator of momentum p emitting a gluon of momentum $-k$. Then we see

$$i\Gamma(p) = \underbrace{(2g)^2 N}_{\text{vertices and index loop}} \underbrace{\int \frac{\mathrm{d}k_-\mathrm{d}k_+}{(2\pi)^2 i}}_{\text{loop integral}} \underbrace{\frac{1}{k_-^2}}_{\text{gluon propagator}} \underbrace{\frac{-i(k_-+p_-)}{(m^2+[2(k_++p_+)-\Gamma(k+p)](k_-+p_-)-i\epsilon)}}_{\text{dressed propagator}}$$

A simple parameter shift $k_+ + p_+ \to k_+$ shows that the right-hand side is independent of p_+ and

[8] One-particle irreducible diagrams; that is, diagrams which cannot be disconnected by cutting one internal edge.

accordingly $\Gamma(p) \equiv \Gamma(p_-)$. Then we have

$$\Gamma(p_-) = \frac{ig^2 N}{\pi^2} \int \frac{\mathrm{d}k_- \, (k_- + p_-)}{k_-^2} \int \frac{\mathrm{d}k_+}{[2k_+ - \Gamma(k_- + p_-)] (k_- + p_-) + m^2 - i\epsilon}$$

where the k_+ integral is fairly trivial except for the fact that it possesses a logarithmic ultraviolet divergence attributed[4] to the light-cone gauge condition. Introducing a cut-off $|k_+| \leq \Lambda$ results in the cancellation of these divergent terms, so we send $\Lambda \to \infty$ and consider this dealt with. Then we can perform a second change of variables in the k_+ integral to now eliminate Γ. This leaves

$$\begin{aligned} \Gamma(p_-) &= \frac{ig^2 N}{\pi^2} \int \frac{\mathrm{d}k_- \, (k_- + p_-)}{k_-^2} \times \frac{+i\pi}{2 \, |k_- + p_-|} \\ &= -\frac{g^2 N}{2\pi} \int \frac{\mathrm{d}k_-}{k_-^2} \mathrm{sign}(k_- + p_-) \end{aligned}$$

with an infrared divergence. To begin with, we take the simplest[5] way of handling this, imposing a *lower* bound on $|k_-|$, as done in 't Hooft's original paper [4].

The cut-off $\lambda < |k_-| < \infty$ gives us

$$\Gamma(p_-) = \frac{g^2 N}{\pi} \left(\frac{1}{|p_-|} - \frac{1}{\lambda} \right) \mathrm{sgn}(p_-)$$

so that the dressed propagator[9] is

$$\frac{-ik_-}{[m^2 - g^2 N/\pi] + k^2 + g^2 N \, |k_-| \, / \pi\lambda - i\epsilon} \tag{3.1}$$

from which we would *like* infer that quark masses are renormalized via

$$m_{(a)}^2 \to M_{(a)}^2 = m_{(a)}^2 - \frac{g^2 N}{\pi} \tag{3.2}$$

However, the cut-off dependent term $\propto \lambda^{-1}$ shifts the pole to infinity, bringing into question exactly what we mean by the *mass* of the quark state. It is very tempting to read this vanishing of (3.1) as being the description of confinement; perhaps arguing that the lack of a pole indicates the absence of a physical state, or equivalently that the quark always has infinite self-energy (away from $k_- = 0$), or that the amplitude for quark propagation vanishes in this limit.

We leave to Appendix B the thorough discussion of a more regular way of addressing the divergence. The prescription discussed there involves fixing the clearly infrared divergent propagator for the massless gluon field by adding a Cauchy principal value symbol P to this propagator. The upshot of the calculation is that equation (3.1) is recovered *without* the $1/\lambda$ term. In this case, we have a *genuine* renormalization $m_{(a)} \to M_{(a)}$ with finite mass quarks (or indeed massless or tachyonic quarks) as given by (3.2). This might look alarmingly contradictory, not just with respect to the formula derived with cut-off, but with our expectation of confinement.

This all raises an issue of some subtlety which is glossed over in 't Hooft's original paper. The key idea is to remember that propagators are *not* amplitudes, but rather parts of Feynman diagrams

[9] For the purposes of drawing comparison with [12] and [4], note that here $g_0^2 = g^2 N$ is the version of the coupling being kept *constant* with respect to the $N \to \infty$ limit. By contrast, [5] uses our notation.

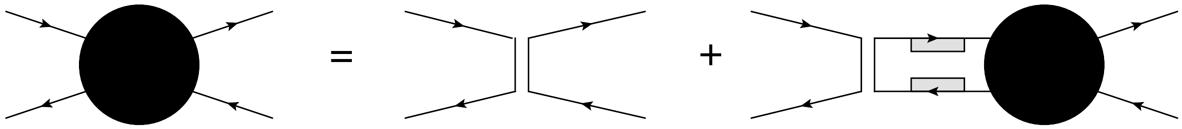


Figure 3.3: Bethe-Salpeter equation for a quark-antiquark scattering amplitude.

allowing the calculation of amplitudes. We must postpone the $\lambda \to 0$ limit of these calculations until they make sense; the dependence on λ should drop out. (We will see this happen in the calculation of the meson spectrum.) In particular, we are compelled to compute only *gauge-invariant* amplitudes. It could be simply asking for the propagator of a free quark of some colour is aphysical regardless of whether free quarks can exist - gauge dependence is restricted to the quark-gluon sector of the full action.

We could instead ask (as 't Hooft did) what the spectrum of a universe consisting of two quarks is, by providing a source for a quark-antiquark pair with no overall colour. If we then discover a *discrete spectrum* of allowed states of some given total momentum p, we infer that there is no possibility of a pair of free quarks propagating separately. This is because we would have to find another continuous label, the *relative* momentum $q = p_1 - p_2$. Instead, we must have only free states of momentum p where the quarks are somehow bound together. These will be *meson* states, which we will now go on to calculate. In Appendix B we discuss how this arises in the regular cut-off prescription.

3.2 Meson States

Now in order to seek out physical quark-based states in our system, we consider scattering amplitudes and look for poles. The simplest non-trivial gauge invariant term we can consider involving quarks is the quark-antiquark amplitude for non-trivial scattering $T_{\alpha\beta,\gamma\delta}(p,p';r)$ (where the Greek indices $\alpha,\beta,\ldots$ are spinor indices which we briefly restore); here p,p' are the incoming and outgoing momenta of the quark, and the total momentum of the system is r. The so-called *Bethe-Salpeter equation* which this satisfies is depicted in Figure 3.3. This translates into the equation[10]

$$T_{\alpha\beta,\gamma\delta}(p,p';r) = \frac{g^2\gamma_{-,\alpha\gamma}\gamma_{-,\beta\delta}}{\left(p_- - p'_-\right)^2} + g^2 N \int \frac{\mathrm{d}k_+\mathrm{d}k_-}{(2\pi)^2 i}\frac{\gamma_{-,\alpha\epsilon}\gamma_{-,\beta\lambda}}{\left(k_- - p_-\right)^2} S_a(k)_{\epsilon\mu} S_b(k-r)_{\lambda\nu} T_{\mu\nu,\gamma\delta}(k,p';r)$$

for $S_a(k)$ the dressed quark propagator for the quark flavour a, up to a γ matrix. We can strip off the γ_- terms to find

$$T(p,p';r) = \frac{g^2}{\left(p_- - p'_-\right)^2} + \frac{4g^2N}{(2\pi)^2 i}\int \frac{\mathrm{d}k_-}{\left(k_- - p_-\right)^2}\phi\left(k_-,p'_-;r\right)$$

where we introduce a new function

$$\phi\left(p_-,p'_-;r\right) = \int \mathrm{d}p_+ \frac{-ip_-}{M_{(a)}^2 + p^2 + \cdots}\cdot\frac{-i\left(p_- - r_-\right)}{M_{(b)}^2 + (p-r)^2 + \cdots} T(p,p';r)$$

and observe that $T(p,\cdots)$ is independent of p_+ and so we can actually do this integral. This is ultimately possible due to the *instantaneous* nature of the interaction mediated by the gluon fields. There are contributions from residues at two poles due to the two propagators which crucially only

[10] This differs slightly from the formula used by Callan et al.[5] simply because we follow the conventions of 't Hooft.

contribute for certain signs of $(r-p)_-$ and p_-; one finds an expression of the form

$$\phi = 2\pi i\,(\cdots)\,T \times \boxed{\theta(p_-)\,\theta(r_- - p_-)}$$

which rearranges to give

$$\left(2\frac{g^2N}{\pi\lambda} + 2r_+ + \frac{M_{(a)}^2}{p_-} + \frac{M_{(b)}^2}{r_- - p_-}\right)\phi\left(p_-, p'_-; r\right) = \theta(p_-)\,\theta(r_- - p_-)$$
$$\times\left[\frac{i\pi g^2}{\left(p_- - p'_-\right)^2} + \frac{g^2N}{\pi}\int \frac{\mathrm{d}\tilde{k}_-}{\left(\tilde{k}_-\right)^2}\phi\left(\tilde{k}_- + p_-, p'_-; r\right)\right]$$

where $\tilde{k}_-$ arises from a change of variables. We note that the $1/\lambda$ term on the left-hand side cancels with that generated by the IR cutoff on the right, where a $2\phi(p_-)/\lambda$ term arises from the integral. Accordingly we drop these terms and interpret the right-hand integral as its principal value[11].

Then letting $x = p_-/r_-, x' = p'_-/r_-$ we find that $\phi \equiv \tilde{\phi}(x, x'; r)$, non-zero only for $x, x' \in [0,1]$. Dropping tildes, and working in this range only,

$$\boxed{\mu^2\phi(x,x';r) = \left(\frac{\alpha_a}{x} + \frac{\alpha_b}{1-x}\right)\phi(x,x';r) - \mathrm{P}\int_0^1 \frac{\mathrm{d}y}{(y-x)^2}\phi(y,x';r) - \frac{i\pi^2}{Nr_-(x-x')^2}} \tag{3.3}$$

where

$$\mu^2 = -\frac{\pi}{g^2N}r^2, \qquad \alpha_a = \frac{\pi M_{(a)}^2}{g^2N} = \frac{\pi m_{(a)}^2}{g^2N} - 1$$

One can solve this equation by dropping the final, $\mathcal{O}(1/N)$ inhomogeneous term in (3.3) and then expanding the actual solution in terms of these eigenfunctions (viewing μ as an eigenvalue). The equation of interest is therefore the simpler

$$\boxed{\mu^2\phi(x,x';r) = H\phi := \left(\frac{\alpha_a}{x} + \frac{\alpha_b}{1-x}\right)\phi(x,x';r) - \mathrm{P}\int_0^1 \frac{\mathrm{d}y}{(y-x)^2}\phi(y,x';r)} \tag{3.4}$$

where we define an operator H acting on the first argument of ϕ. This has roughly the form of a Hamiltonian for a wavefunction; we are effectively seeking a decomposition of a quark-antiquark state into its momentum components.

In Appendix C, we discuss the analytic structure of this equation. The most important fact is that the set of possible μ^2 for which a non-zero solution can exist is *discrete*, giving the crucial result that there is *no continuous spectrum* of quark-antiquark states - that is, there cannot be a pair of asymptotic quarks with independent momenta; instead, there are well-defined *bound states* - and interacting quarks can only scatter into these bound states.

Note that for any bound state arising in this two-particle system, $\mu^2 \sim r^2$ is effectively the squared mass of the quark-antiquark pair. Hence the spectrum of H is essentially telling us the spectrum of mesons in our theory!

Remark. If one develops the interacting theory of mesons discussed in section 3.3 in detail, one can now straightforwardly reconstruct the full scattering amplitude T using the equations in this section; we

[11]That is, we calculate the integral on shifted contours where we replace $\tilde{k}_-$ by $\tilde{k}_- \pm i\epsilon$, then take the mean as $\epsilon \to 0$,

omit this for brevity. Note that the decomposition of the solution (3.3) into the eigenbasis of solutions of (3.4) implies that quark pairs scatter into *superpositions* of the possible meson states, confirming our interpretation. We do not need to do the algebra to deduce this result.

One can go on to ask what the spectrum looks like, particularly for larger mass particles (excited or 'resonance' states, except that in the $N \to \infty$ limit mesons decouple). It is possible to obtain a crude approximation by simply noting that for $\phi(y) \sim \exp(i\omega y)$, the principal value term is dominated by the $y \approx x$ region, and contributes a term $\approx \pi |\omega| \exp(i\omega x)$; thus we expect solutions to exist which are approximately of the form $\exp(i\omega x)$ since this term is compatible with the rest of the equation, at least away from the endpoints. Then our boundary conditions lead to functions like $\phi_k \approx \sqrt{2} \sin k\pi x$, for which $\mu_k^2 \approx \pi^2 k$ if $k \gg 1$. This direct proportionality corresponds to a straight *Regge trajectory*, if these are all the states.

In the case that the bare quark masses $m_{(a)}^2$ were not tachyonic, $\alpha_a \geq -1$ (with equality for massless bare quarks) and so - as may be easily verified by computing $\langle \phi, H\phi \rangle$ using the expression in Appendix C - $\mu^2 \geq 0$. In general, a pair of massless quarks can form a single massless meson with $\phi = 1$, and many other massive states, but if at least one quark is massive then all mesons are massive too.

Remark. In four dimensions, the massless quarks case is of course expected to have a massless Goldstone boson, the 'pion', corresponding to the breaking of chiral symmetry, and so this suggests an interpretation of the massless meson arising in this case. However, Coleman showed[20] that in two dimensions we do not expect Goldstone bosons! Brower[21] resolves this by noting that this meson state decouples from the rest of the theory, as a simple computation from the three-meson vertex discussed in section 3.3 might suggest. See [16] for a discussion of chiral symmetry breaking.

Thus we have successfully solved for the meson spectrum of the theory!

In [4], 't Hooft briefly discusses some numerical results for the masses of bound states which verify (and quantify) the fairly rough analysis provided here. Subsequently, various other standard numerical techniques have been applied, notably the Multhopp transform.[13, 22, 23]

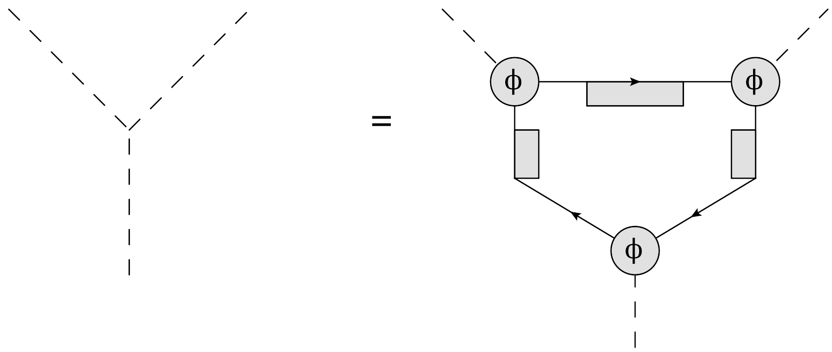
It is straightforward to verify in this way (as [13] details, with illustrations) some perhaps expected results about the resulting wavefunctions. One finds that (a) mesons seem to weigh more than their constituent quarks (with masses in units of $g/\sqrt{2\pi}$, $\mu_{(m_1,m_2)=(1,1)} \approx 2.7$ and $\mu_{(10,10)} \approx 20.55$); (b) heavier quarks seem to give a narrower momentum distribution and a smaller strong force mass contribution $\mu - m_1 - m_2$; (c) the nth state seems to have n stationary points; (d) for unequal quark masses, the heavy quark carries most of the momentum; (e) for light quarks, the wavefunction becomes independent of their masses m since only the gluon field determines most dynamics.

3.3 Effective Meson Theory

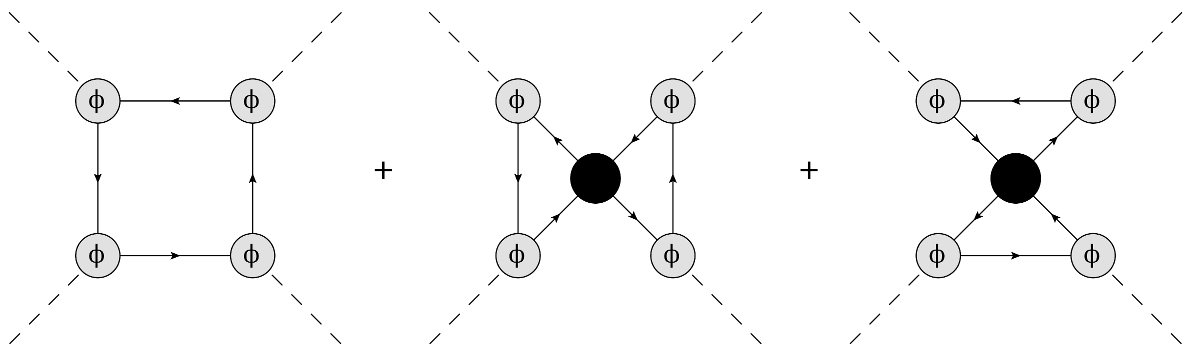
Many other results can be calculated in 't Hooft's framework. In particular, the effective field theory of mesons can be deduced at various orders in $1/N$. Callan et al.[5] for instance calculate the three meson interaction vertex as shown in Figure 3.4a to be

$$\frac{4g^2\sqrt{N}}{\sqrt{\pi}} \Bigg\{ \int_0^{r_-^{(1)}} \mathrm{d}l_- \phi_1\left(\frac{l_-}{r_-^{(1)}}\right) \phi_2\left(\frac{l_-}{r_-^{(2)}}\right) \int_0^{r_-^{(3)}} \mathrm{d}p_- \phi_3\left(\frac{p_-}{r_-^{(3)}}\right) \frac{1}{\left[p_- - \left(l_- - r_-^{(2)}\right)\right]^2}$$

$$- \int_{r_-^{(2)}}^{r_-^{(1)}} \mathrm{d}l_- \phi_1\left(\frac{l_-}{r_-^{(1)}}\right) \phi_3\left(\frac{l_- - r_-^{(2)}}{r_-^{(3)}}\right) \int_0^{r_-^{(2)}} \mathrm{d}p_- \phi_2\left(\frac{p_-}{r_-^{(2)}}\right) \frac{1}{[p_- - l_-]^2} \Bigg\} \quad (3.5)$$



(a) The internal structure of a three-meson vertex



(b) Schematic forms of *some* leading contributions to a four-meson vertex.

Figure 3.4: Diagrams in an effective meson theory

to leading order, where the external momenta are $r_-^{(i)}$; note this scales as $g^2\sqrt{N} = g_0^2/\sqrt{N}$ as observed in section 2.3.[12]

Similarly, the four meson vertices of Figure 3.4b may also be calculated and shown to be finite - however, it is clear that higher order terms become progressively more complicated.

It is important that the results turn out to either be local or involve only the exchange of hadrons. As a result, one concludes that there are *no* long-range forces between mesons (to this order) - this is compared in [5] to the absence of van der Waals forces. The London dispersion force, for instance, has a power law potential $\sim r^{-6}$ due to molecules inducing 'instantaneous' dipoles in each other, even though the molecules are *neutral* in electric charge. Thus neutrality of the bound states is not sufficient to eliminate long-range effects.

A huge amount of calculation (see for instance [21, 24–27]) was done using this model, which constitutes one of the only ways of obtaining exact predictions directly from strongly coupled gauge theories. This aided understanding of the relationships between certain amplitudes. We leave to section 5.3 a general discussion of the value of 1+1 dimensional systems. However, to gain more conceptual insight, we now turn to follow the most suggestive part of the large N approximation: the emergence of the structure of a string theory in the diagrammatic expansion.

[12] Note that [5], like many other papers after 't Hooft, returned to the more physically relevant case of $\mathrm{SU}(N)$ instead of $\mathrm{U}(N)$, which starts to become relevant at higher orders than we have worked with up to this point.

3.4 Quarks on Strings

Before QCD became established as the canonical theory of strong interactions in the 1970s, there were many attempts to build models of the force from 'dual' string theories; with the firm establishment of such gauge field theories, these efforts fell by the wayside to some extent. Yet it seemed there were many features of the strong interaction reminiscent of ideas in string theory, and indeed some agreement of calculations between the theories.

There are various perspectives on the resulting situation: for instance, one can view string theory as an approximate description of the dynamics of quarks and gluons in the confining phase[13] of QCD; alternatively, one could claim that QCD was an effective theory describing some fundamental string theory; or one could hypothesize that in fact the $\mathrm{SU}(3)$ theory is precisely *equivalent* to some dual string theory.

Here QCD, or more generally $\mathrm{SU}(N)$ gauge theory, is thought of as the fundamental theory. In this essay, however, we will see two separate cases for thinking of QCD in terms of strings, motivated by intriguing agreements with $\mathrm{SU}(N)$ theory in the large N limit. Firstly, we will consider an ostensibly very naïve model of mesons as quarks attached to the end of strings and see that it reproduces 't Hooft's equation (3.4), and mention its agreement with various leading order interactions. This may seem to lend itself to an *effective* or phenomenological model of the behaviour of field lines or *flux tubes*. The efficacy of this model is discussed in section 5.3.

In section 4, we will analyse the mathematically richer case of an arbitrary 1+1 dimensional spacetime, in the absence of quarks, where non-trivial degrees of freedom arise due to the non-contractible loops - for the toroidal case, we will see the partition function for QCD looks exactly like that of a string. This illustrates the 'stringiness' of QCD in a more satisfyingly thorough way. Together, we suggest that these viewpoints indicate that strongly interacting gauge theories are cousins of string theory, though we do not claim to have a simple picture of how this correspondence should be made precise. (Though see section 5.3.)

Remark. For a discussion of the modern perspectives on dualities between string theory and field theories, see the review [29]. The AdS/CFT correspondence gives a particular way of viewing certain gauge field theories as string theories, dealing with the particular difficulties of string theories below their *critical dimension* by compactifying many other dimensions. See section 5.2.

In 1976, Bars and Hanson[7] discussed a straightforward, albeit fairly heuristic, method for constructing a theory of strong interactions by placing quarks at the end of strings. We will not dwell much on the underlying logic of their construction, which in fact lies fairly close to the quarks' field theory, instead simply taking their proposed action as described in [8] and following some of their discussion here.

They consider an effective action in d dimensions which is schematically

$$S = \sum_{\text{mesons}} S_{\mathrm{M}} + \sum_{\text{baryons}} S_{\mathrm{B}} + \sum_{\text{quarks}} \left(S^{\text{int}}_{\text{strong}} + S^{\text{int}}_{\mathrm{W}}\right)$$

where we think of a many-body setup rather than a field theoretic one. We will restrict ourselves to the part of the action responsible for the meson spectrum, namely S_{M} (the action for a free meson) neglecting their proposed baryonic model, and electroweak and strong interactions. Importantly, we

[13] One often talks of theories displaying different *phases* at different parameters, characterized by qualitatively different physics. See, for example, [28].

call this *effective* because dynamical gluons (gauge fields) have been completely omitted; colour is now a *global* symmetry. (The authors suggest in [8] that one might pull gauge fields back onto the string worldsheets, but we do not worry about this for our purposes.)

Their model describes a meson as a pair of quarks located at the ends of an open Nambu-Goto string, so

$$S_M = \int \mathrm{d}\tau \left[\mathcal{L}_0 \left(x_1 \left(\tau \right) \right) + \mathcal{L}_0 \left(x_2 \left(\tau \right) \right) - \gamma_M \int_{\text{end } 1}^{\text{end } 2} \mathrm{d}\sigma \sqrt{-h} \right]$$

where $x^\mu \left(\tau, \sigma \right)$ are the string's coordinates, $x_I^\mu \left(\tau \right) \equiv x_I^\mu \left(\tau, \text{end } I \right)$ are the quark coordinates and h is the metric induced on the worldsheet. Here γ_M is the string tension, and the Lagrangians $\mathcal{L}_0$ are spin $1/2$ quark Lagrangians written in terms of the worldsheet.

These Lagrangians are defined for spinor fields with components $\psi_{\alpha I}^{ai} \left(\tau \right)$ - here the index a is a quark flavour index for a mass $m_{(a)}$ quark, i is a colour index, I labels which specific quark we are discussing and α is the spinor index. Suppressing most indices, we take

$$\mathcal{L}_0 \left(x_I \right) = \frac{\partial_\tau x_I^\mu}{2\sqrt{-\left(\partial_\tau x_I \right)^2}} \bar{\psi}_I \gamma_\mu \overleftrightarrow{\partial_\tau} \psi_I - \sqrt{-\left(\partial_\tau x_I \right)^2} \bar{\psi}_I m_{(a)} \psi_I$$

where $a \overleftrightarrow{\partial} b = \left(\partial a \right) b - a \left(\partial b \right)$.

Then the total momentum for S_M is given by

$$\begin{aligned} P^\mu :&= \frac{\delta \mathcal{L}_0 \left(x_1 \right)}{\delta \left(\partial_\tau x_{1,\mu} \right)} + \frac{\delta \mathcal{L}_0 \left(x_2 \right)}{\delta \left(\partial_\tau x_{2,\mu} \right)} - \gamma_M \int_0^\pi \mathrm{d}\sigma \frac{\delta \sqrt{-h}}{\delta \left(\partial_\tau x_\mu \right)} \\ &= p_1^\mu + p_2^\mu + \gamma_M \int_0^\pi \mathrm{d}\sigma \, K^\mu \end{aligned}$$

where in the light-cone gauge, specializing to 1+1 dimensional space, $x^+ \left(\tau, \sigma \right) = \tau$ and then $\left(K^+, K^- \right) = \left(0, \left| \partial_\sigma x^- \right| \right)$. A further ($\partial_\sigma x^-$ uniform) gauge choice gives $K^- = \pi^{-1} \left| x_1^- - x_2^- \right|$. This achieves

$$P^+ = p_1^+ + p_2^+, \qquad P^- = \frac{\tilde{m}_1^2}{2p_1^+} + \frac{\tilde{m}_2^2}{2p_2^+} + \gamma_M \left| x_1^- - x_2^- \right|$$

where $\tilde{m} = \bar{\psi} m \psi = b^\dagger m b + d^\dagger m d$ (after normal ordering) is the mass operator such that $p^2 + \tilde{m}^2 = 0$. Here the operators b, d are defined by

$$\psi_\alpha^{ai} = b^{ai} \left(\tau \right) u_\alpha \left(\tilde{m}, p \left(\tau \right) \right) + d^{\dagger ai} \left(\tau \right) v_\alpha \left(\tilde{m}, p \left(\tau \right) \right), \quad \left(p \cdot \gamma \right) \begin{matrix} u \\ v \end{matrix} = i \tilde{m} \times \begin{matrix} +u \\ -v \end{matrix}$$

and the canonical commutation relations $\left[x_I^-, p_I^+ \right] = -i$, $\left\{ b_I, b_I^\dagger \right\} = 1 = \left\{ d_I, d_I^\dagger \right\}$, with no summation. Note that in this formalism, b, d have no momentum label since the position of ψ is fixed *on the worldsheet*.

At this point, the form of the total momentum P indicates that we are likely to arrive at a very similar result to the 't Hooft result, since the string is clearly simply acting to give a linear Coulomb potential once more. We verify this result for completeness.

Consider two types of normalized state. *Meson states* are colour singlets, of total momentum $P^+ = r^+$ and mass $M^2 = 2P^+ P^- = \mu_n^2$. Also we take *free quark states* where the two quarks are given arbitrary momentum by the choice of $x\left(\tau,\sigma\right)$.

$$\begin{aligned}\text{meson state:}\qquad & \frac{1}{\sqrt{N}}b_1^{\dagger ai}d_2^{\dagger bi}\left|0;n,r^+\right\rangle\\ \text{free quark state:}\qquad & \frac{1}{\sqrt{N}}b_1^{\dagger ai}d_2^{\dagger bi}\left|0;k^+,l^+\right\rangle\end{aligned}$$

Here the '0' refers to the oscillator vacuum. The normalization of meson states is of the form $\left(2\pi\right)\left(2r^+\right)\delta\left(r^+-r'^+\right)\delta_{n,n'}$. The normalization of the free quark states is simply $\delta\left(k^+-k'^+\right)\delta\left(l^+-l'^+\right)$, which is not covariant.

Then we can define wavefunctions for the mesons by

$$\left\langle k^+,l^+\right|\left.n,r^+\right\rangle=2\sqrt{\pi}\delta\left(r^+-k^+-l^+\right)\phi_n^{ab}\left(k^+,l^+\right)$$

(where our gauge choice requires $k^+,l^+>0$) and deduce an equation by considering this same matrix element, but for the operator P^-, using

$$\begin{aligned}P^-\left|n,r^+\right\rangle &= \frac{M^2}{2P^+}\left|n,r^+\right\rangle=\frac{\mu_n^2}{2r^+}\left|n,r^+\right\rangle\\ P^-\left|k^+,l^+\right\rangle &= \left(\frac{\tilde{m}_1^2}{2p_1^+}+\frac{\tilde{m}_2^2}{2p_2^+}+\gamma_M\left|x_1^--x_2^-\right|\right)\left|k^+,l^+\right\rangle\end{aligned}$$

Then we find, via Fourier transform,

$$\begin{aligned}\left(\frac{\mu_n^2}{2r^+}-\frac{m_{(a)}^2}{2k^+}-\frac{m_{(b)}^2}{2l^+}\right)\phi_n^{ab}\left(k^+,l^+\right) &= \gamma_M\int \mathrm{d}k'^+\mathrm{d}l'^+\left\langle k^+,l^+\right|\left(\left|x_1^--x_2^-\right|\right)\left|k'^+,l'^+\right\rangle\phi_n^{ab}\left(k'^+,l'^+\right)\\ &= -\frac{\gamma_M}{\pi}\mathrm{P}\int\frac{\mathrm{d}s^+}{\left(s^+\right)^2}\phi_n^{ab}\left(k^++s^+,l^+-s^+\right)\end{aligned}$$

Now defining the variables $x=k^+/r^+$ and $y=\left(k^++s^+\right)/r^+$ we find ϕ depends only on x by scaling considerations, and the above equation leads to the following, in very close agreement with equation (3.4):

$$\left(\mu_n^2-\frac{m_{(a)}^2}{x}-\frac{m_{(b)}^2}{1-x}\right)\phi_n^{ab}\left(x\right)=-\frac{2\gamma_M}{\pi}\mathrm{P}\int_0^1\frac{\mathrm{d}y}{\left(y-x\right)^2}\phi_n^{ab}\left(y\right)$$

The limits arise due to the positivity constraints on arguments of $\phi\left(k^+,l^+\right)$.

Equivalence with the large N expression[14] (3.4) requires simply that

$$\begin{aligned}2\gamma_M &= g^2N\\ m_{(a)}^2 &= M_{(a)}^2\end{aligned}$$

so that the string tension is essentially the fixed coupling g^2N, and the quark masses are renormalized. (Bars hypothesizes in [8] that the required renormalization occurs when considering his strong interaction term.)

Thus we conclude that the meson spectrum of two dimensional large N QCD is recreated by a theory in which quarks are connected by strings with a tension determined by the fixed coupling

[14] Bars gives a slightly different expression achieved by including a term suppressed by N^{-2}. Also, note that our μ_n here has mass dimensions, whereas in (3.4) we had non-dimensionalized it.

$g_0 \equiv gN^{1/2}$. We point out that the exact form of this tension is slightly special to the two dimensional case on dimensional grounds (i.e. g_0 has a mass dimension here).

Bars goes on to further illustrate that various other quantities, such as the three meson vertex (3.5), are reproduced perturbatively in his model with local quark-quark strong interactions. We will not pursue this analogy further, but consider this suggestive of the existence of a very natural string-like interpretation of the gauge field, at least in this non-self-interacting case.

Specifically, at low energies in the presence of quarks, we expect the gluons to have condensed into a string, with the above tension, linking the quarks. This also explains why we might obtain finite and possibly tachyonic quark propagators in Appendix B when we look at a single quark in the absence of a partner: it is unphysical to consider quarks in a gluon field *without* another point to end the condensed string which supposedly forms between quarks.

4 The 1+1D Case in a Non-Trivial Background

Developing the above hints of string theory, we consider now a more sophisticated development in which the gluon field is treated to all perturbative orders in $1/N$. In 1993, Gross and Taylor[9] - following up on a conjecture of Gross[30] in the spirit of the old dual string line of thought - reinforced the interpretation of gauge theories as strings by means of a comparison of the partition functions for QCD on a two-dimensional manifold and a string theory with this manifold as its target space. They came to an exact equivalence for the case of the torus.

Of course, in a flat spacetime, the 1+1D theory is trivial in the absence of quarks; however, when there are non-contractible loops in the space, then there are 'Wilson lines' which remain as physical degrees of freedom. Intuitively, whilst local bound states of the gauge field cannot exist, excitation modes which are non-contractible waves may exist; these are 'strings of glue'. In their analysis, the partition function due to these loops is compared to that of a weakly coupled string theory term by term in a $1/N$ expansion.

4.1 Partition Functions for Gauge Theories

We begin by observing that we can explicitly calculate the partition function of a pure gauge theory $\mathrm{SU}(N)$ or $\mathrm{U}(N)$ of coupling g for an orientable, two-dimensional manifold $\mathcal{M}$ with genus H, area A and metric h; one finds that

$$\mathcal{Z}_{\mathcal{M}} = \int [\mathrm{D}A^{\mu}] \exp\left(-\frac{1}{4g^2}\int_{\mathcal{M}} \mathrm{d}^2x\,\sqrt{h}\,\mathrm{tr}\left(G^{\mu\nu}G_{\mu\nu}\right)\right)$$

can be evaluated to give[9, 30, 31]

$$\mathcal{Z}_{\mathcal{M}} \equiv \mathcal{Z}\left(H, g_0^2 A, N\right) \stackrel{\mathrm{def}}{=} \sum_R \left(\dim R\right)^{2-2H} e^{-\frac{g_0^2 A}{2N} C_2(R)} \tag{4.1}$$

where the sum is over irreducible representations R of the gauge group, so that $\dim R$ is the dimension and $C_2(R)$ the quadratic Casimir of R. Here $g_0^2 = g^2 N$ exactly as before.

Remark. We will not concern ourselves much with this result, deeming it beyond the scope of the essay; however, it is instructive to see why, for instance, in involves representations of the gauge group. See Appendix D for a brief explanation of this interesting topic. Also, note that the above partition function is obtained from a lattice discretization, where gauge fixing is not required.

Gross conjectured[30] that the free energy $\ln \mathcal{Z}\left(H, g_0^2 A, N\right)$ equals that of a string theory with target space $\mathcal{M}$ where the string coupling is $1/N$ and the string tension is g_0^2. Gross and Taylor demonstrated this with a slightly non-standard formulation of the string theory; rather than a full path integral formulation on this other side, they give a sum over various maps into the target space $\mathcal{M}$ which correspond to a sensible choice of Feynman-like rules for a string theory - that is, a geometrical description is given which coincides with the geometrical description of a (projection onto a specific sector of) string theory. (The suppression of *folds* is the interesting problem, referred to below.)

Since their method is explained in detail in their paper [9], and we are primarily interested in the physical picture and the more conceptual mathematical aspects of their result, we will only sketch their argument here. They begin by attempting to expand (4.1) in the large N limit as an asymptotic series[15] in $1/N$ by expanding each summand separately. This already involves subtleties, since of course the set of representations R changes with N, so it is not obvious what one would mean by 'fixing a representation R' in this limit.

They first discuss the idea of fixing a representation by considering Young diagrams[16] for SU (N) - with rows of length n_i and columns of length c_j, and $n = \sum n_i = \sum c_j$ - and noting the quadratic Casimir is

$$C_2(R) = nN + \underbrace{\left[\sum n_i^2 - \sum c_j^2\right]}_{\tilde{C}(R)} - \frac{n^2}{N}$$

It is arguably better to view (4.1) as a summation over Young tableaux.

Given this, it seems natural to assume that the only representations contributing to (4.1) with a coefficient $\exp\left(-g_0^2 An/2\right)$ would be those with Young tableaux of n boxes. However, they go on to show that the apparently subleading terms in $C_2(R)$ can be made to contribute at leading order by choosing $\mathcal{O}(N)$ length columns in a particular way. Specifically, given two tableaux for representations $R, \tilde{R}$ with column lengths $c_i, \tilde{c}_i$ etc. we construct a 'composite' representation, say $R+\tilde{R}$, with column lengths

$$\begin{cases} N - \tilde{c}_{L+1-i} & i \leq L \\ c_{i-L} & i > L \end{cases}$$

for $L = \tilde{n}_1$. The above Casimir becomes

$$C_2\left(R+\tilde{R}\right) = C_2(R) + C_2\left(\tilde{R}\right) + \frac{2n\tilde{n}}{N}$$

Remark. It will be interesting later to note that the trivial case $n = \tilde{n} = 0$ is exponentially suppressed in the large N limit; that is, $C_2 \sim (n+\tilde{n})N + \mathcal{O}(1)$ so in this trivial case (4.1) implies that there is a non-perturbative $\exp(-\text{const.}/N)$ type suppression. Hence these terms will not appear in our expansion. Gross and Taylor refer to this as the *suppression of folds*; it is equivalent to the absence of the usual variety of stringy particles.

This analysis converts (4.1) into the sum

$$\mathcal{Z} = \sum_n \sum_{\tilde{n}} \sum_{R \in Y_n} \sum_{\tilde{R} \in Y_{\tilde{n}}} \left(\dim\left(R+\tilde{R}\right)\right)^{2(1-H)} \exp\left(-\frac{g_0^2 A}{2N}\left[C_2(R) + C_2\left(\tilde{R}\right) + \frac{2n\tilde{n}}{N}\right]\right) \tag{4.2}$$

where we use their notation Y_n to represent the set of all representations having n squares in their Young tableau. The dimension term has the property that

$$\dim\left(R+\tilde{R}\right) = \dim R \cdot \dim\left(\tilde{R}\right)\left[1 + \mathcal{O}\left(\frac{1}{N^2}\right)\right]$$

[15] As they note, we generically expect QFT calculations which are formal power series expansions in some parameter to be divergent. This is a famous result in QED, for example, though it is generally very easy to understand the reasons for the divergence. In ϕ^4 theory, for example, the four-point coupling λ leads to an obvious divergence in the theory when it is even slightly negative, so any expansion in λ is at best conditionally convergent. Indeed, QCD theory has e^{-N}-type corrections as asymptotic series commonly do - see section 5.1 for how these are related to baryons.

[16] We use the terms *Young tableau* and *Young diagram* interchangeably, though this strictly inaccurate. For an reminder or overview of how these diagrams work, see [32].

which factorizes to order $\mathcal{O}\left(N^{-2}\right)$ just like the rest of $\mathcal{Z}$, where

$$\dim R = \frac{d_R \cdot N^n}{n!} + \mathcal{O}\left(N^{n-1}\right)$$

which is derived in the appendix to [9]. Here, d_R is the dimension of the representation R' of the symmetric group S_n with the *same Young diagram* as R.[17]

Their interpretation of the near-factorization into two summations is that each sum individually takes the form of a 'chiral sector' of a string theory, since each has the form of a sum over coverings of the manifold $\mathcal{M}$ of a fixed orientation (as we will see shortly), so they can be taken to be orientation-preserving or -reversing maps from their domain onto $\mathcal{M}$. As we will sketch below, the coupling term in the *exponential* can also be given an interpretation in this spirit, namely of links or 'tubes' joining together orientation-preserving and orientation-reversing maps. (The $\mathcal{O}(1/N)$ terms in $C_2(R)$ correspond to tubes within a sector.) The authors note that whilst they do not have a stringy interpretation for the subleading terms in the dimension, the torus has genus $H = 1$, and therefore is not sensitive to these at all.

4.2 Geometrical Interpretation

We now wish to offer a geometrical interpretation of the gauge partition function (4.2) in terms of maps between manifolds, with an eye towards a reformulation as a string partition function. Since we are most interested in where the string interpretation arises, we follow the original authors in first considering a single sector, and further omitting the order $1/N$ term in $C_2(R)$.[18] The significance of these terms is explained below.

This setup corresponds to the partition function

$$Z_1 = \sum_{\gamma=-\infty}^{\infty} \sum_{n,i} \zeta_{\gamma,H}^{n,i} \, e^{-\frac{1}{2} n g_0^2 A} \left(g_0^2 A\right)^i N^{2-2\gamma}$$

where

$$\zeta_{\gamma,H}^{n,i} = \begin{cases} \sum_R \left(\frac{n!}{d_R}\right)^{2H-2} \frac{1}{i!} \left(\frac{\tilde{C}(R)}{2}\right)^i & \text{if } 2(\gamma-1) = 2n(H-1)+i \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

Gross and Taylor then demonstrate the correspondence of the sum in Z_1 with a summation over covering maps with the following argument:

- Let M_γ be a space with genus γ, where we allow M_γ to be *disconnected* so that γ (defined by $2-2\gamma = \chi$, the Euler characteristic) may be negative.
- Let $\Sigma(H, n, i)$ be the set of all n-fold covers of M_H (that is, generically n-to-1 maps into $\mathcal{M}$) which have i branch points as their only singularities, where we identify maps up to smooth deformations in the homological sense. (Other point-like singularities will be considered as 'tubes' later.)

[17] Young tableaux are universal in that they describe representations of both symmetric groups and $\mathrm{SU}(N)$; summations over representations R' of S_n and R of $\mathrm{SU}(N)$ are equivalent by just thinking of R as specified by a Young tableau with n boxes. (The reason for this relationship is rooted in the fact that the symmetric group acts on the indices of tensors.) Thus we write summations over R here, and think of it as a summation over Young diagrams as suggested previously.

[18] Looking at the representation theory, this omission can be viewed as making the same $\mathrm{SU}(N) \to \mathrm{U}(N)$ change we made previously.

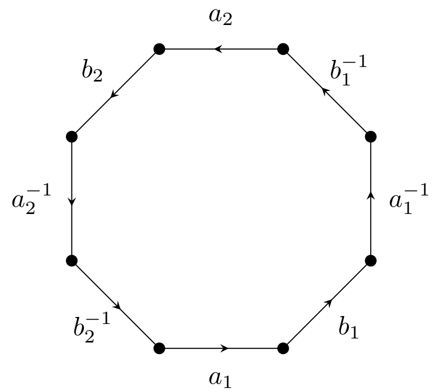


Figure 4.1: Illustration of the construction of the double torus from an octagon. Each vertex is the base point p. The a_1 edge is identified with the a_1^{-1} edge in the opposite direction.

- If $\nu \in \Sigma(H,n,i)$ is a map $\nu : M_\gamma \to \mathcal{M}$ then $2(\gamma-1) = 2n(H-1)+i$. (This can be interpreted geometrically in various ways, for instance: ν winds n times around all holes in $\mathcal{M}$; once around each branch *cut*; but traces completely around M_γ just once including the extra $(n-1)$ covers. Thus $\gamma = nH + i/2 - (n-1)$ as required.)
- Then they demonstrate that

$$\boxed{i!\zeta^{n,i}_{\gamma,H} = \sum_{\nu\in\Sigma(H,n,i)} \frac{1}{|S_\nu|}} \tag{4.4}$$

  where $|S_\nu|$ is a symmetry factor given by the number of distinct homeomorphisms $\mathcal{M}_\gamma \to \mathcal{M}_\gamma$ such that $\nu \circ \pi \equiv \nu$.

In the interests of clarity and brevity, we sketch this last result for the 'unbranched' case $i = 0$ only.[19]

Proof for $i = 0$. We aim to compute the right-hand side of (4.4). Firstly, we seek to understand to understand how the set of possible ν (n-fold covers) can be related to the symmetric group S_n. The ideas, from algebraic topology, are perhaps slightly subtle to grasp. Consider some loop ℓ from a point $p \in \mathcal{M}$ back to itself. Over p there are n *sheets* arising from the n separate times that ν covers $\mathcal{M}$, and hence n points $q_1, \ldots, q_n \in \mathcal{M}_\gamma$ mapping to p under ν. Now one can *lift* the loop ℓ up to the covering space $\mathcal{M}_\gamma$ in n different ways by beginning at each q_k in turn and following a continuous pre-image of ℓ. However, there is no reason for one to end up at the same $q_j \in \mathcal{M}_\gamma$ when one reaches $p \in \mathcal{M}$ by following the original ℓ. That is, following the loop can *permute the n sheets*. This is where the symmetric group arises in considering ν.

To make these notions formal, we consider the fundamental group $\pi_1(\mathcal{M}, p)$, the set of all loops based at some $p \in \mathcal{M}$ identified up to smooth deformations. Then each ν defines a homomorphism $h_{\nu,\text{label}} : \pi_1 \to S_n$ given some labelling. We seek to exchange the summation over ν for a summation over these homomorphisms, noting that all homomorphisms arise via some ν.[34] However, different choices of labelling of the sheets correspond to different h_ν. But $|S_\nu|$ is precisely a measure of this redundancy; for each ν there are $n!/|S_\nu|$ homomorphisms h_ν which can arise from ν.

[19] For an introduction to the fundamental concepts of algebraic topology, and proofs of some of our statements, we refer the reader to [33] as well as the paper [9] itself.

As a result,

$$\sum_{\nu\in\Sigma(H,n,0)}\frac{1}{|S_\nu|}=\sum_h\frac{1}{n!}$$

In order to rewrite this new sum, we consider a set of *generators* for the group $\pi_1(\mathcal{M},p)$, where the group structure involves appending loops to each other. For an H-hole torus, the standard choice is given by a *canonical homology basis* - an informal definition takes loops $a_1,\cdots,a_H$ going around the doughnut holes and $b_1,\cdots,b_H$ going around the corresponding doughnut tube. A more concrete definition is formed by drawing the the $4H$-gon from which $\mathcal{M}$ can be formed by identifying an edge with that two anti-clockwise further round, in the opposite clock direction. All vertices are identified as the base point p. Then going anti-clockwise around the $4H$-gon one traverses the loops $a_1,b_1,a_1^{-1},b_1^{-1},a_2,b_2,a_2^{-1},a_2^{-1},\ldots$ visiting p a total of $4H$ times. See Figure 4.1.

This presentation also makes clear that we should identify a complete tour of the perimeter $a_1b_1a_1^{-1}b_1^{-1}\cdots a_Hb_Ha_H^{-1}b_H^{-1}$ with the trivial loop, since it can be shrunk to a point continuously. In fact, this is the single condition on the group generators needed to restrict the free group to the fundamental group.[33] As a result,

$$\sum_h\frac{1}{n!}=\sum_{s_1,t_1,\ldots,s_H,t_H\in S_n}\frac{1}{n!}\delta\left(s_1t_1s_1^{-1}t_1^{-1}\cdots\right)$$

where $\delta(\rho\in S_n)$ is 1 for $\rho=$ identity and 0 otherwise.

Representation theory of S_n now allows us to find a new expression for this.[20] Given $\rho\in S_n$, $D_R(\rho)=$ matrix representing ρ for Young tableau R and $I_R=D_R(\text{identity})$, the following are standard results:

$$\delta(\rho)=\frac{1}{n!}\sum_R d_R\chi_R(\rho),\qquad\sum_{\rho\in S_n}\chi_R(\rho)D_R\left(\rho^{-1}\right)=\frac{n!}{d_R}I_R,\qquad\sum_{\sigma\in S_n}D_R\left(\sigma\rho\sigma^{-1}\right)=\frac{n!}{d_R}\chi_R(\rho)I_R$$

These imply

$$\begin{aligned}\sum_{s_1,\ldots,t_H}\frac{1}{n!}\delta\left(s_1t_1s_1^{-1}t_1^{-1}\cdots\right)&=\sum_{s_1,\ldots,t_H}\left(\frac{1}{n!}\right)^2\sum_R d_R\chi_R\left(s_1t_1s_1^{-1}t_1^{-1}\cdots\right)\\&=\sum_R d_R\left(\frac{1}{n!}\right)^2\operatorname{tr}\left(\prod_{i=1}^H\sum_{s_i,t_i}D_R\left(s_it_is_i^{-1}\right)D_R\left(t_i^{-1}\right)\right)\\&=\sum_R d_R\left(\frac{1}{n!}\right)^2\operatorname{tr}\left(\prod_{i=1}^H\sum_{t_i}\frac{n!}{d_R}\chi_R(t_i)D_R\left(t_i^{-1}\right)\right)\\&=\sum_R d_R\left(\frac{1}{n!}\right)^2\operatorname{tr}\left(\prod_{i=1}^H\left(\frac{n!}{d_R}\right)^2I_R\right)=\sum_R\left(\frac{n!}{d_R}\right)^{2H-2}\end{aligned}$$

which is in exact agreement with (4.4) by comparison with (4.3) for the $i=0$ case. □

Adding branch points simply complicates the group structure and gives rise to the other terms; the above illustrates the key points of the argument.

[20] Recall we write summations over R here, thinking of them as a summation over Young diagrams.

Let us now recall the omitted $1/N$ term in the expression for $C_2(R)$. As alluded to above, these arise due to 'tubes' between the multiple sheets of a covering map. The idea is that n-covers can fail to be n-to-1 at isolated points; we get point-like singularities where a continuous range gets mapped to a single point. Without worrying too much about the details, we follow the argument of Minahan[35] referred to in [9] and note that factors $\exp\left(ng_0^2 A/2N^2\right)$ can be split into $[n + n(n-1)] \times \cdots$.

Then the first term has the interpretation of arising due to handles in the covering space mapped to points in $\mathcal{M}$. The factors work as follows: there are n sheets; $g_0^2 A$ arises due to the arbitrariness of the location on the sheet; the $1/N^2$ factor arises as we increase the genus needed by one; the $1/2$ factor arises since there is no significance to the orientation of the tube. Similarly, the second term corresponds to the $1/2n(n-1)$ ways of connecting two sheets at a single point of $\mathcal{M}$ (via infinitesimally small 'tubes'). The other factors work out similarly without the indistinguishability. These terms appear in exponentials since one can independently consider contributions to the partition function where one adds as many of these as one wishes.

Finally, we address the question of the cross-terms in (4.2) (for the case of the torus) - specifically the $\exp\left(-g_0^2 A \cdot n\tilde{n}/N^2\right)$ term. In much the same way as the term in $C_2(R)$, these have the interpretation of linking together sheets over $\mathcal{M}$, if we take two sets of maps, one of which preserves orientation and one of which reverses it.[21] The factors are again as they should be: any one of n sheets on one side and $\tilde{n}$ on the other may be joined, at any point on the whole area, increasing the genus by 1, with no symmetry factor. Similarly, the local nature of these modifications leads to an exponentiation. The one slight surprise (see below) is the minus sign here.

4.3 Summary

The final observation left to make is that sums over maps like this are identical to the sort of expansion generated in string theory, if one takes the dictionary given at the beginning of this section:

$$\begin{aligned}
\text{QCD field theory in spacetime } \mathcal{M} &\leftrightarrow \text{QCD string theory with target space } \mathcal{M} \\
\text{number of colours } N &\leftrightarrow \text{string coupling } \frac{1}{N} \\
\text{scaled coupling } g^2 N &\leftrightarrow \text{string tension } g^2 N
\end{aligned}$$

Thus - for the case of the torus $H = 1$ at least - Gross and Taylor establish that the pure QCD theory is perturbatively identical to that of a string theory with *folds* suppressed; for instance, we saw that maps with zero winding number, which necessarily have folds in the string, do not contribute to the stringy partition function. It is a reasonable hypothesis that instability to the formation of thin structures like this is the origin of tachyonic modes in subcritical strings.

They hypothesize that the mechanism in the supposed string action for the absence of these folds may be fermionic fields which cancel their contributions; they also suggest that the unexpected minus sign mentioned above might arise due to fermionic statistics. The presence of such fields could also be invoked as a possible explanation for the $\mathcal{O}(1)$ corrections in the $H \neq 1$ case. All of this is speculative, but seems a reasonable direction for investigation, along with introducing quarks into the theory.

Nonetheless, the conceptual importance of the discussion in this section is well worth emphasizing. We have seen that pure QCD with many colours adopts exactly the form of a correspondingly weakly coupled string theory. It is reasonable to hope that there is a generalization of this duality, since there is no obvious reason for this to be a coincidence. This is the hope expressed in sections 5.2 and 5.3.

[21] Gross and Taylor's paper has a very readable discussion of what a 'tube' connecting orientation-preserving and -reversing sheets looks like.

5 Conclusions

We have seen in this essay that the large N limit gives many insights into the behaviour of non-abelian gauge theories. Let us briefly review some of our key discoveries:

- In section 2 we saw that the large N expansion of QCD is in general highly geometrical, being dominated by diagrams reminiscent of string theory.

- Further, in section 2.3, we saw that if QCD is effectively a theory of mesons, we might expect it to appear as a weakly coupled theory with completely free mesons at $N = \infty$. We also noted that one can explain (simply using geometry and power counting) various phenomena like Zweig's rule and the (leading-order) absence of exotic states.

- In section 3, we found that the possible quark-antiquark states form a discrete spectrum in 1+1 dimensional QCD to leading order. We also verified that these mesons indeed interact sensibly and weakly.

- In section 3.4, we explored the hints of string theory from section 2 by reproducing the spectrum of section 3 using a simple string-like model (and noting that this model can also reproduce meson interaction vertices).

- Finally, in section 4 we considered a much more sophisticated identification of a subtle (toroidal) 1+1 dimensional theory with a string theory to all orders.

In this section we will point to two directions in which our work can be extended - firstly, our attempt to produce some large N hadron phenomenology is clearly limited by the lack of baryons in our theory, so we highlight some ideas and references which explore these ideas. Secondly, we look briefly at where the attempts to use string theories to understand field theories exhibiting string-like behaviour have led - that is, Maldacena's AdS/CFT correspondence and proposed extensions thereof. We will conclude with a discussion of the relevance of 1+1 dimensional theory.

We suggest that we have evidence that the large N model is the an excellent candidate for 'the right approximation' in which to investigate QCD analytically. (Witten also gives an extensive discussion of the attractiveness of this limit in [36].) Of course, the wealth of lattice data provides a more direct numerical way to obtain information. However, such data is difficult to interpret in the absence of any analytical model or approximation; but together they provide an excellent perspective on the theory. We note finally that there is much numerical work confirming that the large N limit described here has sensible properties.[2, 37–40]

5.1 The Problem of Baryons

't Hooft noted in his original paper [3] that

> "It will be clear that in the case of baryons the $1/N$ expansion is extremely delicate."

whilst Coleman said in [12] that

> "For the baryons, things are not so good. Witten's theory is an analytical triumph but a phenomenological disaster."

The reasons for finding difficulty here are clear; it is not clear what a baryon could possibly look like at $N = \infty$ - recall that a baryon is schematically an N-quark state

$$B \sim \epsilon_{i_1 \cdots i_N} q^{i_1} \cdots q^{i_N}$$

This has a mass $\propto N$ at least[22] so that $\exp(ip \cdot x) \propto \exp(-im_{\text{baryon}} t)$ propagators tend to diverge when naïvely expanded as series. This is related to the fact that there are always $\mathcal{O}(N)$ possible ways to attach a gluon line to a quark, so that the number of diagrams diverges with N.

In [36], Witten attempted to address these issues by exploring first non-relativistic quark models of the baryon, and then using a path integral formulation to justify this (Hartree) approximation. He concludes that baryons have a similar status to monopoles in various other theories, which have divergent masses at small 'coupling', here at small $1/\sqrt{N}$. Indeed there are similarities with various instanton solutions, with many amplitudes having a structure like $(1/C)^N$ (making them invisible in perturbation theory) since in $e^+e^- \to B\bar{B}$, say, one must create N quark-antiquark pairs with a multiplicative amplitude C. The classical 't Hooft-Polyakov monopole mass in this sort of theory, for instance, goes as $1/g^2$, though the quantum corrections can be complicated.[41] In this picture, one thinks of the mesons as being analogous to the fundamental fields[23], so that they are indeed subject to $1/\sqrt{N}$ couplings.

Indeed, one can demonstrate (see [42] for instance) that the large N limit unites non-relativistic quark models with soliton models of baryons, such that to some order they give equivalent predictions for some 'model-independent'[43] quantities. For instance, for the ratio of effective couplings[24] $g_{\pi N\Delta}/g_{\pi NN}$, the predictions of the Skyrme model agree with those of large N QCD to order N^2. (This also compares well with experiment - the theoretical predictions are ≈ 1.48 vs. an experimental value of ≈ 1.5.[13, 44])

In short, therefore, it is perfectly possible to obtain sensible results about baryons in this limit by treating them semi-classically as soliton states; the simplest approach fails since one expands in positive powers of N and misses such terms. We see that handling the theory in this qualitatively different way allows us to obtain numerical predictions which are useful in experiment. However, the string interpretation will be more subtle.

5.2 A Nod to AdS/CFT

As remarked in section 3.4, one of the longest frontiers of theoretical physics today is in researching the AdS/CFT correspondence.[29] The reasons are clear - the ability to relate strongly coupled gauge theories to comprehensible gravitational theories to make calculations and predictions tractable hints at the possibility of understanding the dynamics of hitherto deeply mysterious theories.

The basic idea of these theories is seen by viewing the low-energy effective theory of some string theory containing branes in two different ways. For example, the canonical example of the duality between $\text{AdS}_5 \times S^5$ and $\mathcal{N} = 4$ super-Yang-Mills (SYM) theory can be seen by viewing a type IIB string theory in the presence of N nearby 3+1 dimensional branes in a 9+1 dimensional space. At low energies, one finds the massless modes form two decoupled theories: $\mathcal{N} = 4$ SYM theory, with $\text{U}(N)$ gauge group, on the 3+1 branes; and a free supergravity theory in the bulk. But also, one can find a solution of supergravity in the presence of the branes and then investigate what massless modes

[22] Witten takes the N^2 possible terms due to quark-quark interactions to be restrained by a $1/N$ factor in [36].

[23] This is not that unreasonable a viewpoint; the meson spectrum indeed conspires to 'simulate the now-extinct quarks' at large momenta.[5] This corresponds to *asymptotic freedom*, the less immediately interesting phase of QCD, which we did not pursue here.

[24] For instance, $g_{\pi NN}$ is the coupling between the pseudoscalar pion and two nucleons.

propagate along this background - one finds the theory is essentially a free bulk supergravity with some localized excitations close to the origin. Then one identifies the two bulk supergravity solutions with each other; hence one is left with the SYM theory on the one hand, and a type IIB superstring theory in a background which we take to be a local approximation of the supergravity brane solution. Inferring that the modes of these theories may be identified leads to the proposed duality.

For completeness, we sketch how large N QCD may be obtained in the manner of these dualities as described by Aharony et al., in the well-known review [29].

We seek a theory dual to four-dimensional $\mathrm{SU}(N)$ theory; we start from a $(2,0)$ superconformal theory in six spacetime dimensions on N parallel coinciding branes. One compactifies the theory on a circle of radius R_1, obtaining a maximally supersymmetric $\mathrm{SU}(N)$ gauge theory, then on another circle of radius R_0 imposing antiperiodicity conditions on fermions, breaking supersymmetry and giving a dimensionless coupling $g_4^2 = R_1/R_0$. To justify this process, one needs R_0, R_1 to be sufficiently small (compared to $1/\Lambda_{\mathrm{QCD}}$) that the supergravity approximation is invalidated; hence there is work to be done here. Then one makes use of the large N limit to reduce the six-dimensional theory to M theory on $\mathrm{AdS}_7 \times S^4$. The compactification here gives M theory in a black hole background. Then a more standard duality reduces this to a type IIA string theory atop a particular metric $\mathrm{d}s^2 = \cdots$.

The supergravity theory reproduces many of the hoped-for features: most importantly confinement (with a linear potential between colour-charged states) and a mass gap in the glueball spectrum. There is also a large amount of data suggesting that this way of thinking is quantitatively useful too, and we hope that one day a precise formulation of such a duality will allow us to think about gauge theories and string theories almost interchangeably.

5.3 Relevance of 1+1D Models and Strings

There is at first sight something of a deception present in claiming that finding consistently confining solutions in 1+1 dimensions is a justification for expecting confinement in more general theories - specifically, we saw that in this case, the gauge field becomes non-interacting, whilst the linearly increasing nature of the generic Coulomb potential means even Schwinger QED is in many ways a confining theory.[18, 19] That is, it is pure geometry rather than strongly coupled self-interactions which give rise to confinement in these models.

Yet as 't Hooft observes[4]

> "Our result is completely different from the exact solution of two-dimensional massless quantum electrodynamics, which should correspond to $N = 1$ in our case. The perturbation expansion with respect to $1/N$ is then evidently not a good approximation; in two dimensional massless QED the spectrum consists of only one massive particle with the quantum numbers of the photon."

making reference to [19, 45].

To put it another way, we have shown *something* valuable: that these gauge theories, where the low-energy gluon states are one-dimensional in nature, create exactly the conditions we have analysed, giving meson spectra, confinement and string-like structure. But at the same time, we have not shown that real, higher-dimensional theories give rise to these approximately one-dimensional strings of glue.

We assert that we expect to capture some of the relevant physics in this limit. Einhorn et al.[27], for instance, say

> "Experimentally, transverse momenta are observed to be strongly damped in high-energy hadron collisions, so it may even be that certain results obtained in two dimensions

may be abstracted and usefully applied to the real world. Be that as it may, the two-dimensional model poses a well-defined problem wherein the properties of high-energy scattering of bound-state mesons may be evaluated. Obviously, questions related specifically to crossing symmetry, to spin, or to large-transverse-momentum behaviour cannot be approached in two dimensions."

By contrast, Gross and Taylor simply note in their paper[9] that

> "In QCD we expect the physics to be continuous as we vary the dimension between two and four."

In some sense, this is the rather dubious level of physical intuition which is guiding our understanding.

Yet it is plausible that the string-based way of viewing the lower dimensional theories *is* capturing the essential physics of interest:

- The highly suggestive structure of the diagrams and expansion holds in any number of dimensions; we saw this in section 2. It is also discussed at the end of [3].
- There is convincing evidence from numerical lattice calculations of confinement and its 'stringy' nature (the linearity of its potential and hence well-definedness of a phenomenological string tension, which is shown graphically in [2]). See for example [37–40, 46]. The intuitive reason for this is that the effect of gluon self-interaction in higher dimensions is to create a strongly attractive gluon-gluon potential, suppressing virtual gluons (that is, the field strength) away from a stringy excitation - then one obtains a strongly bound pair of quarks in a locally linear world, and a weak coupling between separate strings.

However, in these lattice calculations, often the 'string' length is actually not that much greater than the width - yet in some circumstances still there is a truly *remarkable* agreement[47] with even just the Nambu-Goto spectrum, extending the simple analysis of 3.4, outside the expected radius of convergence. This result is only partially explained by *another* striking observation[48] that any effective string action for an $\mathrm{SU}(N)$ theory *exactly* matches that of the Nambu-Goto action to six-derivative order. These phenomena strongly suggest that some uniting underlying principle (say, a resummation such as that discussed in sections 2 and 3, likely giving a stringy dual, as illustrated in section 4) is controlling the physics of non-abelian gauge theory.

In light of this, perhaps we should reverse the idea that our 1+1 dimensional analyses are irrelevant special cases; maybe the 1+1 dimensional case captures *more* than one initially expects, since the strings which apparently form are 1+1 dimensional objects. (Historically, of course, coarse approximations and special cases have almost always preceded full understanding of a theory.)

The direct extension of the string analysis of section 4 to higher dimensions (further discussed by [9]) gives a different perspective on precisely what is special about the two dimensional case. From this point of view, the reason the two-dimensional string theory (on toroidal spaces at least) was so easy to arrive at in interpreting the QCD expansion is that the extrinsic curvature is trivial in this case. Yet there are reasons to be hopeful; since at least one of the problems inherent in subcritical string theory is absent for the QCD string - we expect the fold-suppressing mechanism in the action removes the *tachyon*. Whatever the nature of this term, we might reasonably expect that the instability leading to the tachyonic modes will be suppressed even in more than two dimensions, since the string is effectively 'rigidified' by this term. This suggests that there is an interesting modification of string theory with physically relevant properties.

Given the lattice evidence[47] and the apparent successes of 'AdS/QCD' duality methods[29], it seems likely that somehow the analysis does generalize, though likely not very straightforwardly. As such, it could be that the 1+1 dimensional case is the only regime in which direct computations are actually comprehensible (or even feasible) - but that it still captures enough of the dynamics to give us a valuable qualitative understanding of the physics involved. Dealing with the many complicating factors of higher dimensional space might be best dealt with in a drastically different way (i.e. via numerics or AdS/QCD), but it might even be possible to explore their effects in terms of the 1+1 dimensional model.

We finish with the suggestion that the viewpoints discussed in this essay give at the very least a much stronger *intuitive* understanding of phenomena in non-abelian gauge theories than can be obtained through any other approach known to the author. Perhaps one day a new Feynman will draw a diagram on the back of an envelope and demonstrate confinement far more simply than any of our current approaches, but for the moment we shall have to make do.

Appendices

A The $\mathrm{U}(1)$ Ghost Field

As stated in section 2, the approximation made in using $\mathrm{U}(N)$ instead of $\mathrm{SU}(N)$ as the gauge group is equivalent to neglecting a $\mathrm{U}(1)_{\text{ghost}}$ ghost gauge field.[25] Since the original generator $\mathrm{U}(1)_{\text{real}}$ which we are attempting to cancel is in the centre of $\mathrm{U}(N)$, it was decoupled from the remaining $\mathrm{SU}(N)$ generators, and hence the $\mathrm{U}(1)_{\text{ghost}}$ field can be taken to commute with the $\mathrm{U}(N)$.

However, this means that the ghost field can only appear as a particle exchanged between quark lines. Another consequence of its not carrying colour charge (clear from the fact it does not interact with the gluons) is that it cannot change the N-counting arguments given. Thus suppose we are given a graph containing some exchanges of ghosts - removing these lines gives a graph of some known order N^s. But reintroducing the ghost lines now suppresses the diagram by some calculable factors: each line brings in a ghost propagator, and two vertices with quark lines. The former, by equation (2.1), must occur with a factor $1/N$. The latter each come with the basic coupling constant g; hence we have an overall relative factor of

$$g^2/N \propto 1/N^2$$

This justifies the observation made in section 2, apart from one interesting subtlety.

Specifically, since we are allowed to remove lines, it is possible that a class of diagrams obtained by removing the ghost is actually larger than that excluding the ghost in the first place. The way this would arise is that purportedly disconnected diagrams could arise from diagrams which are connected by one or more ghost propagators. Then the N suppression of a diagram with a single ghost propagator connecting two quark loops would be reduced by one order relative to the leading diagram (as we have two diagrams of order $N^{2-2H-L} \equiv N^1$ and one ghost with a factor N^{-2} making the diagram $\mathcal{O}(1)$ compared to the $\mathcal{O}(N)$ leading ghost-free quark loop). (Higher number of quark loops remain $1/N^2$ suppressed.)

This is seen not to be an issue, though, since although this diagram cannot arise in $\mathrm{SU}(N)$ theory, due to considerations of colour (or the fact that $\mathrm{tr}T^A = 0$ for $\mathrm{SU}(N)$ generators) the $\mathrm{U}(N)$ theory introduces a diagram with its $\mathrm{U}(1)_{\text{real}}$ boson which is exactly cancelled by the $\mathrm{U}(1)_{\text{ghost}}$ version of the boson.

Accordingly, we generically find that there is a $1/N^2$ suppression of all relevant diagrams involving a ghost field, and hence we can make a reasonably practical theory without the ghost field.

B A Regular Infrared Cutoff

We briefly review arguments of Callan et al.[5] as echoed by Sidney Coleman[12] to see why the divergent quark propagator we found when removing the infrared cut-off is *not* the explanation for quark confinement, and in particular why this does not invalidate our calculations.

The idea is simple: since our prescription in section 3 amounted to separating out the principal value of various integrals from the divergent $1/\lambda$ part, and since we found the $1/\lambda$ terms consistently

[25] See [12, 13] for similar discussions to this.

cancelled, we could simply adopt a regularization procedure in which we do not enforce a $|p_-| \geq \lambda$ cutoff but instead simply *always* work with the principal value of such integrals.

In this version of the theory, there are no infrared divergences whatsoever. This initially sounds like a vast improvement, but as noted in section 3 is surprising, since this leaves us with a perfectly finite quark propagator! We reproduce (3.1) but without the divergent term, so that the mass given by (3.2) is indeed the renormalized mass of an apparent particle. Yet as we explained so long as we only measure *gauge-invariant* quantities, we still recover the discrete spectrum for meson states.

The way this arises in practice is by cancellations of diagrams, though we only see this if we expand sums of diagrams carefully. In the case of the quark-antiquark two-point function, for instance, consider the terms with no gluons in the ladder diagrams. These used to vanish as $\lambda \to 0$ due to the vanishing of the quark propagator, but now are non-zero. This one might expect to result in the appearance of a continuous spectrum. However, a complementary change in the resonance term exactly compensates this new term.

Yet just like the quark propagator, gauge-dependent Green's functions are drastically changed by this change of prescription. This should reinforce the statements in section 3 as to the importance of only calculating gauge-invariant quantities, although this is a reasonable idea anyway.

In essence, what is going on is a *decoupling* of coloured states from the singlets we are interested in. Thus confinement is, from this perspective, a manifestation of a particular decoupling rather than a suppression. Nonetheless, working in the singular case has some advantages since (much like a gauge choice) it makes some calculations simplify, as certain propagators appearing in intermediate calculations vanish in the $\lambda \to 0$ limit.

C The Quark-Antiquark Amplitude

We reproduce equation (3.4) here:

$$\boxed{\mu^2 \phi(x, x'; r) = H\phi := \left(\frac{\alpha_a}{x} + \frac{\alpha_b}{1-x}\right)\phi(x, x'; r) - \mathrm{P}\int_0^1 \frac{\mathrm{d}y}{(y-x)^2}\phi(y, x'; r)}$$

We are very interested in the possible solutions of such an equation; especially in considering for which μ there can be a solution. For this purpose, it is useful to construct the inner product

$$\boxed{\langle \psi, H\phi \rangle = \int_0^1 \left(\frac{\alpha_a + 1}{x} + \frac{\alpha_b + 1}{1-x}\right)\psi^\dagger \phi \mathrm{d}x + \frac{1}{2}\int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \, \frac{(\psi(x) - \psi(y))^\dagger (\phi(x) - \phi(y))}{(x-y)^2}}$$

This is simply defined by $\langle f, g \rangle = \int_0^1 \mathrm{d}x \, f^\dagger g$, but for $\langle \psi, H\phi \rangle$ we simplify the form of the principal value by subtracting off the value of $\phi(x)$ inside the integral in (3.4) and transferring the remainder to the other term; symmetrizing the final integral over x, y gives the expression we use here.

C.1 Boundary Conditions

In order to talk sensibly about the nature of solutions to this equation, we shall consider first the space upon which it most naturally acts, with a view to applying some spectral theory to the eigenvalue problem. Note that near the boundary $x = 0$ the $\alpha_a \phi(x)/x$ term must be balanced by the integral. If we assume (dropping the x' and r dependence for a moment) that $\phi(x) \sim x^\beta$ near $x = 0$, and that the integral is dominated by its lower endpoint, we find

$$\frac{1}{\phi(x)}\mathrm{P}\int_0^1 \frac{\mathrm{d}y}{(y-x)^2}\phi(y) \sim \frac{1}{x}\mathrm{P}\int_0^\infty \frac{\mathrm{d}z}{(z-1)^2}z^\beta = \frac{1}{x}\times -\beta\pi\cot\beta\pi$$

(Note that we expect the error in this approximation to be proportional to $1/\phi(x) = x^{-\beta}$ so for $\beta < 1$ the approximation made in the integral is reasonable.) This implies that

$$\alpha_a + \beta\pi\cot\beta\pi = 0 \tag{C.1}$$

A similar calculation holds at $x = 1$ for α_b. There are infinitely many solutions β to (C.1) for a given α_a, but since $\alpha_a \geq -1$ there is a pair $\pm\beta^\star \in (-1, 1)$ where $\beta^\star > 0$, degenerate for $m_{(a)} = 0$. Then observe that $\beta < -1$ gives a divergent scattering amplitude $T \sim \int \phi$ and that $\beta > 1$ violates our above attempt to balance the most singular terms in (3.4). We conclude that the $\pm\beta^\star$ roots are expected to give the relevant boundary conditions.

It also seems reasonable to expect that $\phi(0), \phi(1)$ vanish, since we know that outside $[0, 1]$ the function ϕ is identically zero. Analysis of the inner product $\langle\psi, H\phi\rangle$ supports this conclusion, since 't Hooft observes[4] that the eigenfunctions of H which vanish at the boundary form a complete set for this space, with other eigenfunctions not orthogonal to these.

Remark. 't Hooft points out that this analysis is not as rigorous as one might hope; without considering at least $1/N$ corrections, it is difficult to be sure that we are constructing amplitudes compatible with the principles of field theory, particularly unitarity (conservation of probability).

C.2 Discreteness and Completeness of Spectrum

We give a quick argument as to why we expect the discreteness of the solutions to $H\phi = \lambda\phi$ for H a linear, self-adjoint operator. Suppose that one has a *continuous family* of solutions $[H\phi(\cdot;\lambda)](x) = \lambda\phi(x;\lambda)$ with solutions depending smoothly on λ. Then differentiate with respect to λ, and assume that the linearity of H allows us to commute ∂_λ through it. Hence $H\partial_\lambda\phi = \lambda\partial_\lambda\phi + \phi$. Then take the inner product with ϕ:

$$\langle\phi, H\partial_\lambda\phi\rangle = \lambda\langle\phi, \partial_\lambda\phi\rangle + \langle\phi, \phi\rangle$$

But now if we assume $\partial_\lambda\phi$ is in the space of solutions upon which H is self-adjoint (which is the case for our homogeneous boundary conditions on a compact interval) then we can rewrite this as

$$\langle\phi, \phi\rangle = \langle H\phi, \partial_\lambda\phi\rangle - \lambda\langle\phi, \partial_\lambda\phi\rangle = -2\mathrm{Im}(\lambda)\langle\phi, \partial_\lambda\phi\rangle = 0$$

since $\lambda \in \mathbb{R}$. Thus such a family of solutions is identically zero.

To prove completeness of the eigenfunctions, the usual spectral theory approach is to establish that the discrete set of eigenvalues form a sequence $\lambda_n \to \infty$ with orthonormal eigenfunctions u_n, and then to bound the error in estimating f formed by projecting out the $u_1, \ldots, u_n$ parts of a function f - one finds this goes as const.$/\lambda_{n+1} \to 0$.

D A Perspective on Gauge Theoretic Partition Functions

The analysis carried out by Migdal[31] approaches the computation of the partition function via renormalization, effectively from a lattice regularization. One computes various quantities for regions in the decomposition of the manifold - in particular what Rusakov calls the K-functional K_p for a 'plaquette' - which are *gauge-invariant* functions which one takes to depend only on path-ordered exponentials knows as *Wilson loops* evaluated near this site. (The highly geometrical nature of the above expression for the partition function is partly tied up with the fact that expectations of Wilson loops are intimately tied up with the area they enclose.) This allows one to apply some moderately sophisticated harmonic analysis, essentially representation theory, to simplify the integrations, in what is essentially a generalization of the Fourier transform. (In finite lattice theories with compact gauge groups, there is not the usual divergence due to the infinite volume of the total group of all gauge symmetries, so there is no need for a gauge-fixing process such as that implemented in the first part of this essay.)

Specifically, we expand a function on the gauge group

$$K_p\left(U_p\right) := \exp\left(\beta_0 N \cdot \operatorname{tr}\left(U_p + U_p^\dagger\right)\right) \overset{!}{=} \sum_R \dim R \cdot \lambda_R \, \chi_R\left(U_p\right)$$

where the coefficients $\lambda_R = \lambda_R\left(\beta_0 N\right)$ are determined using the orthogonality relation

$$\int \mathrm{d}U \, \chi_R\left(U_1 U\right) \chi_{R'}\left(U^\dagger U_2\right) = \frac{\chi_R\left(U_1 U_2\right) \delta_{R,R'}}{\dim R}$$

(or rather the special case $U_1 = U_2 = I$ so that the right-hand side reduces to simply $\delta_{R,R'}$) and where $\chi_R\left(U\right)$ is the character (trace) of the matrix U in the representation R. The Haar measure of group integration is assumed to be chosen to give the group unit volume. Further, the fact that $\lambda_R\left(z\right) \sim 1 - C_2\left(R\right)/2z + \cdots$ as $z \to \infty$ (and that $\beta_0 \propto 1/a^2 \to 0$ where the lattice spacing $a \to 0$) gives rise to the appearance of the quadratic Casimir in these expressions.

Hence the appearance of the sum over representations - formally an application of harmonic analysis and results such as the Peter-Weyl theorem - is best thought of as arising due to the different (representation) components functions on the gauge group can have. This is exactly analogous to the way in which the representations $e^{-ip\cdot x}$ of the translation group form a basis for expanding functions on the real line.

Finally, we note that in section 5 of [9], the authors describe a different method of arriving at the final results of section 4 more in the spirit of this appendix.

References

[1] D. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973), no. 26 1343–1346.

[2] G. S. Bali, *The Running Coupling from Lattice QCD*, in *Prodvino 1993, Problems on high energy physics and field theory*, pp. 147–163, 1993.

[3] G. 't Hooft, *A planar diagram theory for strong interactions*, *Nucl. Phys. B* **B72** (1974) 461–473.

[4] G. 't Hooft, *A two-dimensional model for mesons*, *Nucl. Phys. B* **B75** (1974) 461–470.

[5] C. G. Callan Jr., N. Coote, and D. J. Gross, *Two-dimensional Yang-Mills theory: A model of quark confinement*, *Phys. Rev. D: Part. Fields* **13** (1976), no. 6 1649–1669.

[6] I. Bars and M. B. Green, *Poincaré- and gauge-invariant two-dimensional quantum chromodynamics*, *Phys. Rev. D* **17** (1978) 537–545.

[7] I. Bars and A. J. Hanson, *Quarks at the ends of the string*, *Phys. Rev. D: Part. Fields* **13** (1976), no. 6 1744–1760.

[8] I. Bars, *A quantum string theory of hadrons and its relation to quantum chromodynamics in two dimensions*, *Nucl. Phys. B* **111** (1976) 413–440.

[9] D. Gross and W. Taylor, *Two-dimensional QCD is a string theory*, *Nucl. Phys. B* **400** (1993), no. 1-3 181–208.

[10] D. Amati and M. Testa, *Quark imprisonment as the origin of strong interactions*, *Phys. Lett. B* **48** (1974), no. 3 227–231.

[11] G. Rajasekaran and V. Srinivasan, *Generation of gluons from quark confinement*, *Pramana* **10** (1978), no. 1 33–45.

[12] S. Coleman, *Aspects of Symmetry.* Cambridge University Press, 1988.

[13] A. Manohar, *Large N QCD*, in *Probing the Standard Model of Particle Interactions (Les Houches Summer School)* (F. David and R. Gupta, eds.), 2004.

[14] M. A. Luty, *Baryons with many colors and flavors*, *Phys. Rev. D* **51** (1995) 2322–2331.

[15] P. Kovtun, M. Ünsal, and L. G. Yaffe, *Volume independence in large N_c QCD-like gauge theories*, *JHEP* **2007** (2007), no. 06 019.

[16] S. Coleman and E. Witten, *Chiral-Symmetry Breakdown in Large-N Chromodynamics*, *Phys. Rev. Lett.* **45** (1980), no. 2 100–102.

[17] V. Višnjić-Triantafillou, *Zweig rule in two-dimensional QCD*, *Nucl. Phys. B* **164** (1980) 495–508.

[18] S. R. Coleman, *More About the Massive Schwinger Model*, *Ann. Phys.* **101** (1976) 239–267.

[19] J. Schwinger, *Gauge Invariance and Mass. II*, *Phys. Rev.* **128** (1962), no. 5 2425–2429.

[20] S. Coleman, *There are no Goldstone bosons in two dimensions*, *Commun. Math. Phys.* **31** (1973), no. 4 259–264.

[21] R. Brower, J. Ellis, M. Schmidt, and J. Weis, *Hadron scattering in two-dimensional QCD: (I). Formalism and leading order calculations*, *Nucl. Phys. B* **128** (1977), no. 1 131–174.

[22] B. Grinstein and R. F. Lebed, *Explicit quark-hadron duality in heavy-light meson weak decays in the 't Hooft model*, *Phys. Rev. D* **57** (1998), no. 3 1366–1378.

[23] R. Jaffe and P. F. Mende, *When is field theory effective?*, *Nucl. Phys. B* **369** (1992), no. 1-2 189–218.

[24] R. Brower, J. Ellis, M. Schmidt, and J. Weis, *Hadron scattering in two-dimensional QCD: (II). Second-order calculations, multi-Regge and inclusive reactions*, *Nucl. Phys. B* **128** (1977), no. 1 175–203.

[25] J. Cardy, *Quark confinement in two dimensions and the parton model*, *Phys. Lett. B* **61** (1976), no. 3 293–296.

[26] M. B. Einhorn, *Confinement, form factors, and deep-inelastic scattering in two-dimensional quantum chromodynamics*, *Phys. Rev. D* **14** (1976), no. 12 3451–3471.

[27] M. B. Einhorn, S. Nussinov, and E. Rabinovici, *Meson scattering in quantum chromodynamics in two dimensions*, *Phys. Rev. D* **15** (1977), no. 8 2282–2286.

[28] C. Ratti, M. A. Thaler, and W. Weise, *Phases of QCD: Lattice thermodynamics and a field theoretical model*, *Phys. Rev. D* **73** (2006), no. 1.

[29] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, *Large N field theories, string theory and gravity*, *Physics Reports* **323** (2000), no. 3-4 183–386.

[30] D. J. Gross, *Two-dimensional QCD as a string theory*, *Nucl. Phys. B* **400** (1993), no. 1-3 161–180.

[31] A. Migdal, *Recursion equations in gauge field theories*, *Sov. Phys. JETP.* **42** (1976) 413.

[32] M. Vaughn, *Introduction to Mathematical Physics.* Wiley, 2008.

[33] A. Hatcher, *Algebraic Topology.* Cambridge University Press, 2002.

[34] H. Zieschang, E. Vogt, and H.-D. Coldewey, *Surfaces and planar discontinuous groups, Lecture Notes in Mathematics*, vol. 835. Springer-Verlag, 1980.

[35] J. A. Minahan, *Summing over inequivalent maps in the string theory interpretation of two-dimensional QCD*, *Phys. Rev. D* **47** (1993) 3430–3436.

[36] E. Witten, *Baryons in the* $1/N$ *expansion*, *Nucl. Phys. B* **160** (1979), no. 1 57–115.

[37] M. Teper, *Large* N_C *physics from the lattice*, in *Phenomenology of large* N_C *QCD. Proceedings* (R. Lebed, ed.), 2002.

[38] M. Panero, *Recent results in large-N lattice gauge theories*, in *PoS (Lattice 2012)* (D. Leinweber, ed.), 2012.

[39] B. Lucini and M. Panero, *SU(N) gauge theories at large N*, *preprint* (2012). arXiv:1210.4997 [hep-th].

[40] G. Bali, L. Castagnini, S. Collins, F. Bursa, L. D. Debbio, B. Lucini, and M. Panero, *Meson spectrum in large N QCQ*, in *PoS (Confinement X)*, no. 278, 2012.

[41] A. Rajantie, *Mass of a quantum 't Hooft-Polyakov monopole*, *JHEP* **2006** (2006), no. 01 088.

[42] A. V. Manohar, *Equivalence of the chiral soliton and quark models in large* N, *Nucl. Phys. B* **248** (1984), no. 1 19–28.

[43] G. S. Adkins and C. R. Nappi, *Model independent relations for baryons as solitons in mesonic theories*, *Nucl. Phys. B* **249** (1985), no. 3 507–518.

[44] G. S. Adkins, C. R. Nappi, and E. Witten, *Static properties of nucleons in the Skyrme model*, *Nucl. Phys. B* **228** (1983), no. 3 552–566.

[45] A. Casher, J. Kogut, and L. Susskind, *Vacuum Polarization and the Quark-Parton Puzzle*, *Phys. Rev. Lett.* **31** (1973) 792–795.

[46] M. Lüscher and P. Weisz, *Quark confinement and the bosonic string*, *JHEP* **2002** (2002), no. 07 049.

[47] A. Athenodorou, B. Bringoltz, and M. Teper, *The spectrum of closed loops of fundamental flux in D = 3+1* $\mathrm{SU}(N)$ *gauge theories*, in *PoS (Lattice 2009)*, no. 223, 2009.

[48] O. Aharony and E. Karzbrun, *On the effective action of confining strings*, *JHEP* **2009** (2009), no. 06 012.