PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Gauge Interactions, Elementarity and Superunification

Abdus Salam

Phil. Trans. R. Soc. Lond. A 1982 **304**, 135-153
doi: 10.1098/rsta.1982.0010

Phil. Trans. R. Soc. Lond. **304**, 135–153 (1982) [135]
Printed in Great Britain

Gauge interactions, elementarity and superunification

By ABDUS SALAM, F.R.S.

International Centre for Theoretical Physics, P.O.B. 586, *Miramare-Strada Costiera* 11, 34100 *Trieste, Italy and Department of Physics, The Blackett Laboratory, Imperial College, London SW*7 2*BZ, U.K.*

Present thinking on the concept of elementarity of particles and forces is reviewed as are the ideas behind superunification of gravity and the electronuclear force.

1. Introduction

Particle physics, as we know it today, began some eighty-five years ago with J.J. Thomson's discovery of the electron, and Lorentz's bold extrapolation of Maxwell's electrodynamics down to the distances of the electron's 'classical' radius. Assuming that the 'family' concept currently used to classify particles is correct, the identification of the companions of the electron, essentially constituting the first family, took around forty years of experimentation, as did the identification of the strong and the weak nuclear forces governing their mutual interactions. The second family began with the cosmic-ray discovery of the muon and required also forty years for its completion.

Contrast this relatively slow development, ranging over more than seventy-five years, with the revolutionary changes registered by the subject during the last decade. Not only was the second family completed and a third nearly so, but, more important, the experimental work during the decade, made possible by availability of detection devices and higher accelerator energies, gave us confidence in the essential correctness of gauge ideas – the subject of this symposium – for describing *elementary forces*. The first result of this has been the pushing up of the energy frontier, over which it now appears possible to ask meaningful questions, from a few GeV to Planck energies of the order of 10^{19} GeV, with a corresponding pushing back of the time frontier from 10^{-7} s to 10^{-44} s, within the context of a big-bang model of the early Universe. A second result has been the possible obliteration of the traditional distinction between electromagnetic, nuclear and gravitational forces.

The greatness of gauge ideas for phenomenological physics lies in the circumstance that, through their use, two of the basic questions of (i) what are the elementary constituents of matter and (ii) what are the elementary forces among them? get interrelated with each other through the concept of *elementary charges*. Describing elementary particles as the basic carriers of certain elementary charges (gravitational, electrical and nuclear) one finds that the gauge forces turn out (at the first approximation) to be proportional to these charges. A postulated symmetry among the charges, then, leads directly to a unification hypothesis among the elementary forces.†

This is important, but the real import of gauge theories goes deeper. The elementary charges mentioned above – and the field-theoretic currents associated with them – are rooted, according to our present ideas, within the symmetries of space and time, and the symmetries of mysterious manifolds describing the internal structure of elementary particles. By focusing on these

† Gauge theories, besides their role in describing and motivating a unification of elementary forces, have also revealed the possible existence of rich topological structures, e.g. instantons and monopoles.

symmetries, gauge theories provide us with windows on topological (and other) structures of space and time as well as of the internal manifolds and appear to motivate an intimate synthesis between them.

A part of the package of these symmetry ideas is the study of the observed patterns of symmetry-breaking and in particular the breaking of symmetries spontaneously. *Spontaneous* symmetry-breaking has the character of a transition phenomenon, with the possibility of symmetry restoration, revealed in suitable environments of temperature, space–time curvature, topology, or external electric and magnetic fields. An important part of our study relates to the energies – the mass scales – where such transitions occur.†

The subject has thus been transformed during the last decade through the twin studies of gauge symmetries and their spontaneous breaking. But, these advances notwithstanding, we are still very far from the elucidation of the nature of the elementary charges or of the problems posed by the mass scales. In this paper my first task is to consider, in the light of the gauge ideas, the question: Is the very concept of elementarity, of charges, forces and particles tied to the mass scale? My second task is more specific: to consider a possible unification of the gravitational with the electronuclear force near Planck energies (*ca.* 10^{19} GeV), through a gauging of a newly discovered – and before 1971 wholly unsuspected – symmetry between bosons and fermions, called supersymmetry. This is the superunification in the title of this paper, which not only promises to achieve a unified theory of matter and its interactions, but also attempts to find a geometrical meaning for the elementary charges it uses within *extended* space–time.

2. Two perspectives on elementarity and electronuclear gauge theories

2.1. *The concept of elementarity*

Consider first the concept of elementarity for particles and gauges. At least as far as the *electronuclear* phenomena are concerned, two points of view are at present expressed:

(*a*) We have discovered the ultimate elementary particles; they are the quarks and the leptons, represented by a renormalizable gauge theory effective over all energies, with no length parameter in the interaction, i.e. the field theoretic 'radius' of the particles is zero. Intermediate mass scales, whose origin is obscure, are introduced as (Higgs) parameters in the Lagrangian. As energy increases, beyond successive intermediate mass scales, the symmetry of the theory also progressively increases.

(*b*) A contrasting point of view states that gauge symmetries are not golden calves to be worshipped; that there are stages of elementarity dependent on the energy; that quarks and leptons are composed of prequarks (preons), preons are possibly composites of pre-preons, pre-preons of pre-pre-preons, ... At each energy stage effective Lagrangians exist. The symmetries relevant to effective Lagrangians for the light composites may differ in different energy régimes; in fact symmetries may even decrease as energy increases. The intermediate mass scales may correspond to the different levels of elementarity.

† The transition phenomena associated with the onset of spontaneous symmetry-breaking and its restoration, at higher mass scales – as revealed by cosmological remnants of epochs gone by – have knit particle physics and cosmology more intimately together.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

2.2. *The first view: grand-unifying theories*

The first point of view is exemplified by the grand-unifying theories (GUTs) and I shall briefly review these, emphasizing in particular the intermediate mass scales, and the possibility that *all* symmetry-breaking phenomena are spontaneously realized.

(*a*) The preceding papers give evidence that low energy phenomena exhibit exact

$$SU_C(3)\times U_{E.M.}(1)$$

symmetries of chromodynamics and quantum electrodynamics. Quantum chromodynamics (QCD) is a remarkable theory. Besides asymptotic freedom, i.e. a decreasing coupling parameter α_s for increasing energy ($\alpha_s(q^2)\rightarrow(12\pi/33)(\ln q^2/\Lambda_c^2)^{-1}$ for $q^2\gg\Lambda_c^2$), the theory is believed to confine exactly, i.e. colour symmetry is an invisible symmetry in the physical spectrum.

(*b*) The preceding papers also give evidence, as energy increases beyond 100 GeV, for the $U_{E.M.}(1)$ symmetry to increase to $U_{L,R}(1)\times SU_L(2)$ of the electroweak force. At this stage the expected symmetry group $SU_C(3)\times U_{L,R}(1)\times SU_L(2)$ is characterized by *three* independent parameters α_s, α and $\sin^2\theta$ (or alternatively Λ_c, α, $\sin^2\theta$ if α_s is dimensionally transmuted in favour of the mass parameter Λ_c). The standard model of three families of quarks and leptons, with 15 two-component particles in each family, uses a Higgs doublet to generate spontaneous symmetry-breaking of $SU_L(2)\times U_{L,R}(1)$ to $U_{E.M.}(1)$ for energies below about 100 GeV. The model needs at least 26 empirically determined parameters for its specification – a daunting task for the eventual theory.

(*c*) For energies above about 250 GeV the symmetry represented by $U_{L,R}(1)$ may expand into $U_{B-L}(1)\times SU_R(2)$, connoting a left–right symmetry. An extension is expected in each family from 15 two-component quarks and leptons to 16 (i.e. each ν_L is accompanied by ν_R), and new gauge bosons $W_R^{\pm}$ coupling with V + A currents.

(*d*) Increasing energies† beyond 10^4–10^5 GeV may increase the symmetry

$$[SU_C(3)\times U_{B-L}(1)]\times[SU_R(2)\times SU_L(2)]\quad\text{to}\quad SU_C(4)\times SU_R(2)\times SU_L(2).$$

The four colour symmetry $SU_C(4)$ (Pati & Salam 1974), which may supersede

$$[SU_C(3)\times U_{B-L}(1)]$$

beyond 10^4–10^5 GeV, would be the first master symmetry exhibiting a fundamental quark–lepton unification, in the sense that quarks as well as leptons would be described as members of one irreducible multiplet of a single symmetry group $SU_C(4)$. The left–right symmetric, $SU_C(4)\times SU_R(2)\times SU_L(2)$, would depend on *two* coupling parameters α_s and $\alpha\sin^{-2}\theta$. This is also the first unification stage where (on account of the non-Abelian character of the groups concerned) all charges must appear quantized. The spontaneously broken $SU_C(4)$ permits proton decays (Pati & Salam 1973 *b*) into three leptons (for example $P\rightarrow 3\nu+\pi^+$, $N\rightarrow e^-+2\nu+\pi^+$) as well as neutron–antineutron oscillations at the level of $10^7\lesssim\tau_{N-\bar{N}}\lesssim10^8$ s, the level to which experiments are currently directed.

† Note the vast separation between the expected succession of mass scales: *ca.* 250 GeV, 10^4–10^5 GeV, ... Even with the promise of technicolour with its characteristic mass scale of about 1000 GeV, we are entering the age either of a true passimony of nature for new phenomena, or of our theoretical bankruptcy in recognizing important empirical clues. Clearly we desperately need new experimental inputs.

(*e*) The next step of grand unification, the increasing of symmetry such that the theory registers just one gauge constant, may come about in three ways.

(i) The 'flavour' symmetries ($SU_L(2) \times SU_R(2)$) for any one family may become part of a flavour $SU_F(4)$, with multiplets containing also mirror-quarks and mirror-leptons. These need be no heavier than about 300 GeV. A discrete flavour–colour symmetry between $SU_C(4)$ and $SU_F(4)$ would then ensure one coupling parameter for a grand unifying symmetry

$$SU_C(4) \times SU_F(4)$$

emerging beyond 10^{14} GeV. In this model, appropriate Higgs multiplets could bring about proton decays in the mode $P \rightarrow \bar{l}$ (for example $P \rightarrow e^+ + \pi^0$, $N \rightarrow e^+ + \pi^-$).

TABLE 1

	allowed processes	appropriate mass scales/GeV
$SU(16) \rightarrow SU(8) \times SU(8) \times U_{3B+L}(1)$	$P \rightarrow e^+\pi^0$	10^{14}
$\rightarrow SU(2) \times SU(2) \times SU(4)$	$\rightarrow e^-\pi^+\pi^+$	10^9–10^{10}
$\rightarrow$	$\rightarrow e^-\nu\nu\pi^+\pi^+$	10^4–10^5
	$N \rightarrow \bar{N}$	not allowed
$SU(16) \rightarrow SO(10)$	$P \rightarrow e^+\pi^0$	10^{14}
$\rightarrow SU(2) \times SU(2) \times SU(4)$	$N \rightarrow \bar{N}$	not allowed
$SU(16) \rightarrow SU_q(12) \times SU_l(4) \times U(1)$	$P \rightarrow e^+\pi^0$	10^{14}
$\rightarrow SU(2) \times SU(2) \times SU(3) \times U(1)$	$N \rightarrow \bar{N}$	10^4–10^5

(ii) The symmetry $SU_C(4) \times SU_R(2) \times SU_L(2)$ may be part of an SO(10) that manifests itself for energies in excess of 10^{14} GeV. The 16-fold spinor multiplet of SO(10) would contain left-handed quarks and leptons as well as left-handed anti-quarks and anti-leptons. (Alternatively there may be no intermediate $SU_C(4) \times SU_R(2) \times SU_L(2)$ stage; the $SU_C(3) \times SU(2) \times U(1)$ may expand directly into SU(5) for energies exceeding 10^{14} GeV. This is the energy (Λ_0) for which the three couplings (for $SU_C(3)$, $SU_L(2)$ and U(1)) converge to a common value, when renormalization group techniques are used (Georgi *et al.* 1974). The assumptions that go into the computation of Λ_0 are: there are no new forces up to Λ_0 (including forces that might differentiate between the three families), or new particles (which might upset $\sin^2\theta(\Lambda_0)$ from its value of $\frac{3}{8}$, derived on the basis of known particles). Thus if we *assume* that there exists a desert of new phenomena up to Λ_0, renormalization group techniques then tell us that Λ_0 is high, *ca.* 10^{14} GeV. Note that so long as a grand-unifying group G descends with one mass scale Λ_0 down to $SU(3) \times SU(2) \times U(1)$, and so long as $\sin^2\theta(\Lambda_0) = \frac{3}{8}$, *any* G (e.g. SU(5), SO(10), SU(16)) will give identical predictions for proton decay.) It may be difficult to accommodate in the minimal SU(5) and SO(10) models, without introducing extra intermediate mass scales through extra Higgs multiplets, N–$\bar{N}$ oscillations at the present level of experimentation (assuming these are discovered), or a ν_e-mass of around 10 eV (assuming this is confirmed), or proton decay into a lepton and pions

$$(P \rightarrow e^- + \pi^+ + \pi^+).$$

(iii) Finally there is the possibility of the maximal gauge symmetry being realized: this is SU(16), the maximal symmetry that could hold for 16 two-component quarks and leptons and their antiparticles) belonging to one family (Pati *et al.* 1975 *a*, *b*, 1980 *b*, *c*). (Here again anomaly cancellation makes mirror-particles mandatory.) This symmetry could permit the coexistence of four types of decay modes for the proton; $P \rightarrow l^+$, $P \rightarrow l^-$, $P \rightarrow 3l$, $P \rightarrow 3\bar{l}$ alongside N–$\bar{N}$

oscillations, at the currently planned level of experimentation. There will be three intermediate mass scales: *ca.* 10^{14} GeV, 10^9–10^{10} GeV, and 10^4–10^5 GeV (see table 1).

The important difference between SU(16) and SO(10), so far as proton decays of the type $P \to e^+ + \pi^0$ are concerned, lies in the circumstance that the decay is intrinsic in the SO(10) (or SU(5)) model. Even beyond 10^{14} GeV, when the gauge particles concerned may have undergone a transition to masslessness, the decay $P \to uud \to e^+$ continues unabated. For the SU(16) model, on the other hand, where $P \to e^+\pi^0$ is a consequence of spontaneous symmetry-breaking, the transition $P \to uud \to e^+$ will cease when symmetry is restored beyond 10^{14} GeV. Of course SU(16) permits, in addition, decays of the type $P \to e^-\pi^+\pi^+$, $P \to 3\nu + \pi^+$, $P \to 3\bar{\nu} + \pi^+$, forbidden, for example, in minimal SU(5). (Presumably, after this stage (of SU(16) or its siblings) would come the uniting of families into 'tribes'. I do not discuss this.)

2.3. *The second view: elementarity and gauges*

A contrasting view to GUTs posits that there is no linear progression of increasing symmetry as energy increases; that intermediate mass scales do exist but that they represent new levels of elementarity. Quarks and leptons are composites, made of pre-quarks (preons); preons may be composites of pre-preons; pre-preons of pre-pre-preons and so on (Pati *et al.* 1975*a*, *b*, 1980*b*, *c*; Pati *et al.* 1980*a* and references therein).

This view has surfaced because of discontent with the following.

(*a*) Far too many Higgs multiplets are needed in grand unified theories like SO(10) or SU(16), being necessary if several intermediate mass scales exist.

(*b*) There are far too many quarks and leptons (39 two-component ones already discovered at the last count; six more awaited) for them to qualify as an elementary set.

(*c*) There are too many gauge bosons; symmetry groups are too large.

(*d*) Finally, intermediate mass scales are too widely spaced (technically 'unnatural'), for example 10^2 GeV and 10^{14} GeV in minimal SU(5).

The simplest preonic model (Curtright & Freund 1979) with quarks and leptons as composites of preons assumes eight preons, f_u, f_d; C_R, C_Y, C_B; F_1, F_2, F_3: two flavons f_u, f_d carrying flavour, three chromons C_R, C_Y, C_B carrying colour, and three familons F_1, F_2, F_3. The light preons would correspond to an SU(8) symmetry containing $SU(5) \times SU_{\text{family}}(3)$.

An alternative that illustrates the notion of differing symmetries at the composite (quark–lepton) level compared with the symmetries of preonic theory is Harari & Seiberg's (1981) model. There are 18 preons, Tohu's (T) and Vohu's (V), with intrinsic symmetry

$$SU_C(3) \times SU_{\text{h.c.}}(3) \times [U(1)]^2, \quad T_{L,R} = (3,3)_{L,R}, \quad V_{L,R} = (3,\bar{3})_{L,R}.$$

Here $SU_{\text{h.c.}}(3)$ is a hypercolour group with an appropriate $\Lambda_{\text{h.c.}}$ in the TeV range. (Though Harari & Seiberg do not take this point of view, one may, if one is economy conscious, consider the 18 preons themselves as composites of six pre-preons (C_R, C_Y, C_B; H_R, H_Y, H_B).) Quarks and leptons, singlets of hypercolour, are TTT, TTV, TVV and VVV composites. As energy *decreases* below $\Lambda_{\text{h.c.}}$, i.e. at the composite (q, l) level, the symmetry group is not $SU_C(3) \times SU_{\text{h.c.}}(3) \times [U(1)]^2$ but $SU_C(3) \times SU_L(2) \times SU_R(2) \times U_{B-L}(1)$, with the implication that $W_L^{\pm}$, $W_R^{\pm}$ are *composite* gauges; the corresponding charges are non-elementary. The $W^{\pm}$ forces are Van-der-Waals forces between hypercolour neutral composite objects.

The following question now arises: why does the weak Van-der-Waals force (mediated by the

$W^{\pm}$) exhibit such an elegant Yang–Mills character? A second and related question is: why are quarks and leptons (composites of preons) so light compared with $\Lambda_{\mathrm{h.c.}}$? What symmetry protects them against acquiring (heavy) masses?

A lore has developed in answer to both these questions, tied to chirality as the protector of fermions† against mass-acquisition, and renormalizability as protector of spin-1 particles against non-Yang–Mills behaviour. A chiral spinor is massless; a vector meson theory with no associated mass scale is renormalizable if it is Yang–Mills and vice versa. Let us envisage the following scenario. Let there be a succession of colour-like theories, colour, hypercolour, hyper-hypercolour, ..., with associated mass scales $\Lambda_{\mathrm{c}}, \Lambda_{\mathrm{h.c.}}, \Lambda_{\mathrm{h.h.c.}}, \ldots$. Why there are these mass scales, and how to determine one in terms of the next remain a mystery. Assume the theory

Table 2

energy régime	'elementary entities'	light composites
$\Lambda_{\mathrm{c}} \leftrightarrow \Lambda_{\mathrm{h.c.}}$	(q, l)	hadrons; singlets of colour
$\Lambda_{\mathrm{h.c.}} \leftrightarrow \Lambda_{\mathrm{h.h.c.}}$	preons	(q, l); singlets of hypercolour
$\Lambda_{\mathrm{h.h.c.}} \leftrightarrow \Lambda_{\mathrm{h.h.h.c.}}$	pre-preons	preons; singlets of hyper-hypercolour
...	...	...

derives all other masses dynamically from these. Let quarks and leptons be 'elementary' below $\Lambda_{\mathrm{h.c.}}$. Quantitatively within the energy range Λ_{c}–$\Lambda_{\mathrm{h.c.}}$, let these describe all physical phenomena through an 'elementary' Lagrangian. Preons play this role in the range $\Lambda_{\mathrm{h.c.}} \leftrightarrow \Lambda_{\mathrm{h.h.c.}}$; pre-preons in the range $\Lambda_{\mathrm{h.h.c.}} \leftrightarrow \Lambda_{\mathrm{h.h.h.c.}}$; .... The lengths $\Lambda_{\mathrm{c}}^{-1}, \Lambda_{\mathrm{h.c.}}^{-1}, \Lambda_{\mathrm{h.h.c.}}^{-1}, \ldots$ are the confinement (bag) radii of the singlet light composites of colour (hadrons), hypercolour (quarks and leptons), hyper-hypercolour (preons), ... Thus we have the picture given in table 2. (It could be that leptons (or at least e and ν_e) are 'elementary entities' of a level different from that of quarks, and their role is only that of spectators (for anomaly cancellations) for the régime indicated in table 2.) If in any energy régime, the physics can be described through a renormalizable field theory of 'elementary entities' or, equally, through an effective Lagrangian of fields corresponding to composites (both light and heavy) made out of the 'elementary entities' of the preceeding energy régime (decoupling theorem), then the following *ansatz* should hold (Veltman, as quoted in Ellis *et al.* 1980*b*) up to energies of around $\Lambda_{\mathrm{n}+1}$:

$$\begin{aligned}\mathscr{L}^{\text{renormalizable}}_{\text{elementary}} \approx \mathscr{L}^{\text{renormalizable}}_{\text{effective}} &\text{ (light composites)}\\ &+\Lambda_{\mathrm{n}}^{-1}\,\mathscr{L}^{\text{non-renormalizable}}_{\text{effective}} \text{ (light and heavy composites)}\\ &+\text{terms of order } \Lambda_{\mathrm{n}}^{-2}+\ldots.\end{aligned}$$

This *ansatz*‡ makes it plausible that:

(*a*) Light composites are spin-zero, spin-$\frac{1}{2}$, or spin-1; if spin-1, they must be Yang–Mills. Renormalizability imposes freedom from anomalies.

(*b*) Composites of spin-$\frac{3}{2}$, spin-2 or higher must be heavy, generating non-renormalizable interactions with $\mathscr{L}_{\text{effective}}$ damped by powers of $\Lambda_{\mathrm{n}}^{-1}$.

(*c*) $\mathscr{L}^{\text{renormalizable}}_{\text{elementary}}$ and $\mathscr{L}^{\text{renormalizable}}_{\text{effective}}$ may exhibit different (gauge) symmetries.

Which are the possible light composites for a given 'elementary' theory? Clearly combi-

† As shown in other papers given at the symposium, in the context of supertechnicolour, supersymmetry for fermions and bosons can protect scalar companions of chiral fermions from mass acquisition.

‡ The terms on the right-hand side of the *ansatz* also give the limitations of $\mathscr{L}^{\text{renormalizable}}_{\text{effective}}$ (light composites); this piece by itself may give a fair approximation up to Λ_{n} only.

natorics of spins and other symmetries must play a role, but the real problem is one of dynamics. 't Hooft (1979), however, has sought to reduce it to a group-theoretic problem. His major tool is chiral invariance and matching of chiral anomalies for the elementary and the renormalizable effective Lagrangian, together with certain decoupling requirements. I do not discuss 't Hooft's criteria, or their implementation by Banks *et al.* (1980) and Bars & Yankelowicz (1981), except to remark that (surprisingly) the implementation involves representations of graded algebras.

Before concluding this section let me remark that Ellis, Gaillard, Maiani & Zumino (1980*a*; hereafter E.G.M.Z.) have considered a preon supergravity model which I shall be discussing later. Anticipating, however, the elementarity and mass scales they might propose, we have:

mass scale	Λ_C ⟷	$\Lambda_{GUT}(10^{14}\,\text{GeV})$ ⟷	Planck (10^{19} GeV)
elementary entities	(q, l)		preons
symmetry group	$SU(5)\times SU_{\text{family}}(3)$		SU(8)

Beyond 10^{19} GeV, Pati *et al.* (1980*b*, *c*) suggest that there may be a pre-preonic régime, of (analogue) electric and magnetic dyons. The symmetry group is the humble $U_E(1)\times U_M(1)$, with gravity itself an induced phenomenon. We now consider gravity and its unification with the electronuclear force.

3. Unification of electronuclear forces with gravity; eleven space–time dimensions

A vast extrapolation (some sixty orders of magnitude) is required if we are to believe that Einstein's gravity theory, with its dimensional constant $\kappa = (16\pi G_N)^{\frac{1}{2}} \approx (10^{19}\,\text{GeV})^{-1}$, devised originally to describe long-range phenomena, will continue to hold down to distances $\kappa \approx 10^{-33}$ cm. If such an extrapolation makes sense, the presence of the dimensional constant κ in Einstein's theory clearly sets it apart from the electronuclear theory where the gauge coupling is dimensionless. A unification of the electronuclear and gravitational theories must be construed in the sense that Einstein envisaged: the electronuclear charges must find a niche in the geometry of space–time, like the gravitational charge which in Einstein's theory found an association with the geometrical notion of space–time curvature.

This type of unification was accomplished, in a remarkable theory, between Maxwell's electromagnetism and gravity by Kaluza (1921) and subsequently Klein (1926). They suggested that the electric charge may be identified with the fifth component of momentum in a space–time extending to five dimensions. Formally Kaluza showed that the scalar curvature in a five-dimensional space–time equals Einstein's Lagrangian (which is scalar curvature in four dimensions) plus Maxwell's Lagrangian, in standard interaction with gravity, provided the electromagnetic potential is identified with the $g_{5\mu}$ component of the metric. More specifically, by writing,

$$g_{MN} = \left|\begin{array}{c|c} g_{\mu\nu}-\kappa^2 A_\mu A_\nu & \kappa A_\mu \\ \hline \kappa A_\nu & -1 \end{array}\right|, \quad M, N = 0, 1, 2, 3, 5, \quad \mu, \nu = 0, 1, 2, 3,$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

the action

$$S = -\frac{1}{4\kappa^2}\int_0^L \frac{dx^5}{L}\int d^4x g^{\frac{1}{2}}_{MN} R_5$$

equals the sum of the standard Einstein and Maxwell actions if $g_{\mu\nu}$ and A_μ are independent of the fifth coordinate x_5.

Two types of objection were raised against this unification. Einstein objected: he could not see how other matter, and particularly spinor matter, could be geometrical. We shall see later that this objection is met today through supersymmetry which unites fermions with bosons. The second objection came from Pauli (1933): electricity and gravity had separated like oil and water in this theory. Surely somewhere there ought to be new testable consequences of the unification suggested. One might indeed discern new consequences for charged spin-$\frac{1}{2}$ particles, but these appeared physically disastrous, at least in 1933.

To see this, following Thirring (1972), one can write the Lagrangian for a spin-$\frac{1}{2}$ fermion in five dimensions in the form,†

$$i\overline{\psi}[\gamma^\mu(\partial_\mu + ieA_\mu) + M]\psi + \tfrac{1}{16}i\kappa\overline{\psi}(\gamma^5 m_0/M - i\mu/M)F_{\mu\nu}[\gamma^\mu, \gamma^\nu]\psi$$

where, for the dependence on the fifth coordinate x_5, we have assumed that

$$\psi(x, x_5) = \exp(i\mu x_5)\psi(x).$$

Here $M = (m_0^2 + \mu^2)^{\frac{1}{2}}$, while $\kappa\mu$ is the electric charge e. Note that:

(*a*) We have assumed that the five-dimensional manifold is a product, $M^4 \times S^1$, of the four-dimensional Minkowski manifold M^4 with the circle S^1 of size $2\pi/\mu = 2\pi\kappa/e$. The fifth dimension has thus curled up to a size of the order of 10^{-33} cm.

(*b*) The charged fermion mass M is *ca.* 10^{19} GeV.

(*c*) The charged particle carries a non-zero electric dipole moment, which violates both P and T.

Since 1933 we have become used, not only to T-violation,‡ but also to particles of Planck mass, Pauli's objections do not have the same force today.

I shall not pursue fermions any further in connection with Kaluza–Klein theory, since we shall subsequently be formulating supergravity theories that contain fermions in extended space–times. Here, I merely cite a recent paper by Witten (1981) which considers the problem of finding the manifold of minimum dimensionality that could support unification of

$$SU_C(3) \times SU_L(2) \times U(1)$$

with Einstein's gravity, in the Kaluza–Klein sense.

Let B be the internal space parametrized by $\phi^i, i = 1, \ldots, n$, for an internal symmetry G (symmetry generators $T^a, a = 1, \ldots, N$). Write the generalized Kaluza–Klein metric in the form

† Note the natural emergence of moment terms, with their conserved topological currents $\partial_\mu(\overline{\psi}\sigma_{\mu\nu}\psi)$, and $\partial_\mu(\overline{\psi}\gamma_5\sigma_{\mu\nu}\psi)$. The corresponding (commuting) charges are central charges, in a supersymmetric context (see §4.1 and Salam & Strathdee (1978), Salam (1978)).

‡ Thirring (1972) suggested that to push up the magnitude of T-violation to the level observed in K-phenomena, one needs to consider seriously a spin-2 Kaluza–Klein theory of strong *f*-gravity (cf. Isham *et al.* 1971). I must emphasize here the virtues of strongly interacting composites of spin-2 made up, for example, of 'elementary' gluons or 'hyper-gluons' or 'hyper-hyper-gluons', ... Since the only ghost-free spin-2 equation known is the Einstein equation, the fields representing such composites must satisfy it. Further, these composites could acquire induced mass terms that are dynamically generated through a mechanism analogous to gluon condensation, and are described by Salam & Strathdee (1976). The chief virtue of these spin-2 composites is that they would confine quarks or gluons within bags of the size of their inverse masses (Salam & Strathdee 1978, Baaklini & Salam 1979).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

$$g_{AB}(x^\alpha, \phi^k) = \left|\begin{array}{c|c} g_{\mu\nu}(x^\alpha) & \sum_a A^a_\nu(x^\alpha)\, K^a_j(\phi^k) \\ \hline \sum_a A^a_\mu(x^\alpha)\, K^a_i(\phi^k) & \gamma_{ij}(\phi^k) \end{array}\right|$$

Here γ_{ij} is the metric of the internal space B, $A^a_\mu(x^\alpha)$ are massless gauge fields of G, and K^a_i are the appropriate Killing vectors.

What is the manifold B of minimal dimensionality that can support

$$SU(3) \times SU(2) \times U(1)?$$

Now U(1) is the symmetry group of the circle S^1, with dimension one. The space of lowest dimension with symmetry SU(2) is the sphere S^2 with dimension two, while the space of lowest dimension with symmetry SU(3) is the complex projective space CP^2 with dimension four. Thus the space $CP^2 \times S^2 \times S^1$ can support $SU(3) \times SU(2) \times U(1)$ and has dimension $4+2+1 = 7$. With four non-compact 'space–time' dimensions, the total dimensionality of our world must then be $4+7 = 11$, if gravity as well as $SU(3) \times SU(2) \times U(1)$ are to be supported in a gauge fashion according to the ideas of Kaluza and Klein.

This is a remarkable result. We shall see later, that there is a maximum of eleven dimensions for supergravity. This is because supergravities in higher dimensions probably contain massless particles of spins greater than two. The existence of such particles would contradict many of the fundamental assumptions of the quantum theory of fields.

We may indeed have been living in eleven-dimensional space–time all the time but no one knew till 1979 when $SU(3) \times SU(2) \times U(1)$ symmetry was first clearly established! Eleven, as a number, has the merit, that to my knowledge, nothing mystical has ever been associated with it.

In the next section supergravity theories are discussed, with their twin unifications of gravity with the electronuclear force and fermions with bosons.

4. Supersymmetry, supergravity and superunification

4.1. *Supersymmetry, simple and extended*

Supersymmetry is the symmetry between fermions and bosons. That for a simple Bose–Fermi free theory,

$$\mathscr{L} = -\tfrac{1}{2}(\partial_\mu A)^2 - \tfrac{1}{2}(\partial_\mu B) - \tfrac{1}{2}\bar{\lambda}\gamma\,\partial\lambda,$$

one can 'rotate' bosons into fermions came as a profound surprise when Golfand & Lichtman (1971) (and, independently, Wess & Zumino (1973)) discovered this remarkable symmetry. (The rotation is in a spinor-space extension of space–time:

$$\delta A = \tfrac{1}{2}\bar{\epsilon}\lambda, \quad \delta B = -\tfrac{1}{2}\epsilon\gamma_5\lambda, \quad \delta\lambda = \tfrac{1}{2}(A - i\gamma_5 B)\,\epsilon.$$

Here A and B are massless spin-zero, and λ is a Majorana field; ϵ is the constant infinitesimal spinor parameter of rotation. The Lagrangian $\mathscr{L}$ is invariant up to a divergence. The symmetry algebra closes on-shell, i.e. one must use equations of motion to demonstrate closure, though by adjoining two auxiliary non-propagating fields to the physical set (A, B, λ), one can secure closure even off-shell for this theory. This is not always possible, however, and poses one of the unsolved problems in the subject: on-shell supersymmetric Lagrangians are often available, but not their off-shell counterparts.)

Turning to interactions, one may deduce, for example, the Lagrangian for a Higgs (spin-zero) field from a Lagrangian for spin-$\frac{1}{2}$ quarks. Likewise gauge bosons (spin 1) and gravitons (spin 2) would be accompanied by and interact with gauge fermions (gauginos) and spin-$\frac{3}{2}$ gravitinos, respectively, in a supersymmetric theory. (A supersymmetric theory of gravity would thus realize Einstein's dream of elevating the 'base wood' of (fermion) matter on the right-hand side of his equation $R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R = -T_{\mu\nu}$, to the status of (spin-2 bosonic) 'marble' of gravity on the left-hand side. As we shall see, this dream has come to be realized in a manner Einstein would have approved: not only can supersymmetric gravity be formulated; but it is also *the gauge theory of supersymmetry*.)

Since supersymmetry transformations convert bosons into fermions, the supersymmetry generator, the supersymmetry charge Q_α, must be a (Majorana) spinor. One can demonstrate the anti-commutator relation:

$$\{Q_\alpha, \bar{Q}_\beta\} = 2(\gamma_\mu)_{\alpha\beta}P^\mu.$$

Here P_μ is the energy–momentum vector. Clearly supersymmetry represents an extension of Poincaré's space–time symmetry.

Exceedingly important for physical applications are the *N-extended supersymmetries* introduced by Salam & Strathdee (1974) where the supersymmetric generators $Q^i_\alpha (i = 1, 2, ..., N)$, correspond to the fundamental representation of an internal symmetry SO(N). The Q^i_α satisfy†

$$\{Q^i_\alpha, \bar{Q}^j_\beta\} = 2(\gamma_\mu)_{\alpha\beta}P^\mu\delta^{ij},$$

$$[Q^i_\alpha, P_\mu] = 0.$$

For the simple $(N = 1)$ supersymmetry, the massless super-multiplets consist of helicity states $(\pm\frac{1}{2}, 0^+, 0^-)$, $(\pm 1, \pm\frac{1}{2})$, $(\pm\frac{3}{2}, \pm 1)$, or $(\pm 2, \pm\frac{3}{2})$; for $N = 2$, the helicity content of the fundamental supermultiplet is $(\pm 1, \mathbf{2}\times(\pm\frac{1}{2}), 2\times 0)$; for $N = 4$ the helicity content is $(\pm 1, \mathbf{4}\times(\pm\frac{1}{2});$ $\mathbf{6}\times 0)$. Here $\mathbf{4}\times(\pm\frac{1}{2})$ (for example) stands for *four* massless Majorana fields, representing $\mathbf{4}\times 2$ physical degrees of freedom, making up a fourfold of SO(4); likewise the six spin-zero objects represent a sixfold multiplet of SO(4).

Extended supersymmetries corresponding to $N = 2$ and $N = 4$ are particularly interesting: the content of the supermultiplets is the same as of (compactified) $N = 1$ simple supersymmetry supermultiplets in six- and ten-dimensional space–time respectively. (Consider, for example, $N = 1$ simple supersymmetry in ten dimensions, represented by a ten-component field A_P and a 16-component spinor ψ_R. After compactification down to four dimensions, the 10-vector A_P appears as a 4-vector A_μ plus six scalars $A_5, A_6, ..., A_{10}$. Likewise the 16-component spinor $\psi_R \equiv \psi_{\alpha i}(\alpha, i = 1, ..., 4)$ has the content of four Majorana spinors in four-dimensional space–time. This parentage of the $N = 4$ extended supersymmetry from ten dimensions (SO(9, 1) $\supset$ SO(3, 1) $\times$ SO(6)) anticipates that a hidden 'internal' symmetry exists and that it is likely to be as large as SO(6) $\approx$ SU(4) rather than just SO(4).)

† More generally $\{Q^i_\alpha, \bar{Q}^j_\beta\} = 2(\gamma_\mu)_{\alpha\beta}P^\mu\delta^{ij} + Z^{ij}\delta_{\alpha\beta} + (\gamma_5)_{\alpha\beta}Z'^{ij}$. Here the $\frac{1}{2}n(n-1)$ Z^{ij} and $\frac{1}{2}n(n-1)$ Z'^{ij} are the so-called central charges which, on-shell and in flat space-time, commute with each other and with P_μ.

4.2. $N = 4$ *Yang–Mills theory and its finiteness*

A renormalizable on-shell supersymmetric Lagrangian for the $N = 4$ multiplet can be written down. If we introduce external non-Abelian local symmetry G, say $SU(k)$, such that the total symmetry is $SU(k) \times (N = 4$, supersymmetry), we would be dealing with a Yang–Mills theory with $k^2 - 1$ fields of helicity ± 1, $4(k^2-1)$ of helicity $\pm\frac{1}{2}$ and $6(k^2-1)$ of helicity zero, with a unique coupling parameter g. A remarkable property then holds. It has been verified by direct calculations that for up to three loops the renormalization-group β-function ($\beta(g) = M\,\mathrm{d}g/\mathrm{d}M$) vanishes identically. If this result could be proved generally, for all loops, this would be the first *finite* infinity-free quantum field theory in physics. (The Green functions of this theory may exhibit gauge-dependent infinities. Such infinities are inconsequential, since in the supersymmetric analogue of the axial gauge even these would be absent, if $\beta(g)$ vanishes.)

Is this theory really finite to all orders?

A general proof for this miracle has been given by Sohnius & West (1981) and by Ferrara & Zumino (unpublished). The proof relies on several assumptions, plausible but not all demonstrated in a *supersymmetric* context. The proof relies on: (i) the identity $\Theta^{\mu}_{\mu} = [\beta(g)/2g^3]\,F^i_{\mu\nu}F^i_{\mu\nu}$, (ii) the absence of all anomalies, (iii) the structure of the supersymmetric conformal anomaly-multiplet and (iv) the hidden SU(4) 'internal' symmetry exhibited by this particular theory, which has been mentioned already.

For the one-loop case an alternative proof has been given by Curtright (1981). The Yang–Mills current can be split into convective and magnetic parts. Their contributions to the β-function are

$$\beta = M\frac{\mathrm{d}g}{\mathrm{d}M} = \frac{hC}{96\pi^2}\sum_S (1-12S^2)(-1)^{2S}$$

where C is the appropriate quadratic invariant for the gauge group representation carried by the particle, and the sum is over the moduli (S) of the helicities. From the helicity content of the $N = 4$ theory discussed above, it is easy to verify that

$$\sum_S 1(-1)^{2S} = 0 = \sum_S S^2(-1)^{2S}.$$

Thus the convective and the magnetic infinities vanish individually.†

† Each time the 'magnetic' Yang–Mills coupling acts, one may expect an additional factor of S in the expression for β in the Curtright formula. If this is so, the l-loop generalization of β may read: $\beta = a_0 + a_1 s^2 + \ldots + a_l s^{2l}$, and the individual vanishing of convective and magnetic infinities may not carry over to more than one loop. Such a cancellation was a consequence of the following identities for extended supermultiplets:

$$\sum_j (-1)^{2S(j)}[S(j)]^k D(j) = 0, \quad k = 0, 1, \ldots, N-1,$$
$$\sum_j (-1)^{2S(j)}[S(j)]^k C(j) = 0, \quad k = 0, 1, \ldots, N-3,$$
$$\sum_j (-1)^{2S(j)}[S(j)]^k F(j) = 0, \quad k = 0, 1, \ldots, N-5,$$

Here $D(j)$, $C(j)$ and $F(j)$, are the dimension, and the quadratic and quartic invariants for the 'internal' SO(N) (or the hidden 'internal' SU(N)) symmetry group associated with N-extended supersymmetry, at the helicity $S(j)$. Clearly for $N = 4$, these formulae cannot take us beyond one loop. For two and three loops. either there are other sources of infinity cancellation or, as has sometimes been the case with infinities, we are following a red herring. This view is bound to be unpopular but there appear to be some conjectural reasons why the fourth loop may be infinite.

4.3. *Extended supergravities*

Supergravity with $N = 1$ is the gauge theory of simple supersymmetry in the same sense that Einstein's gravity is the gauge theory of the Lorentz–Poincaré symmetry. We motivate this, by noting that for $N = 1$ supersymmetry, the anticommutator for charges

$$\{Q_\alpha, \bar{Q}_\beta\} = 2(\gamma_\mu)_{\alpha\beta} P^\mu$$

may be expected to generalize to

$$\{Q_\alpha, J_{\beta\lambda}(x)\} = 2(\gamma_\mu)_{\alpha\beta} \Theta^\mu_\lambda(x)$$

where $J_{\beta\lambda}(x)$ is the current corresponding to the charge Q_β, and $\Theta^\mu_\lambda(x)$ the energy–momentum tensor is the current of P_μ. Clearly a supersymmetric generalization of the gauge theory that associates a spin-2 graviton with $\Theta^\mu_\lambda(x)$ must be a theory that also associates a gravitino of spin-$\frac{3}{2}$ with the supersymmetry current $J_{\beta\lambda}(x)$.

Such a theory was formulated by Freedman *et al.* (1976) and independently by Deser & Zumino (1976). The Lagrangian reads

$$\mathscr{L} = \mathscr{L}_{\text{Einstein}}(e, \omega) + \mathscr{L}_{\text{Rarita–Schwinger}}(\psi_\mu, e, \omega)$$

where e, ω, ψ are the vierbein, spin connection and spin-$\frac{3}{2}$ Rarita–Schwinger fields, with the covariant derivative D_ρ and ω defined as

$$\mathrm{D}_\rho = \partial_\rho + \tfrac{1}{2}\omega_\rho^{mn}\sigma_{mn},$$

$$\omega_\mu^{mn} = \omega_\mu^{mn}(e) + \tfrac{1}{4}(\bar{\psi}_\mu \gamma^m \psi^n - n \leftrightarrow m + \bar{\psi}^m \gamma_\mu \psi^n).$$

Of fundamental importance for a programme of superunification of gravity with the electronuclear force are the extended supergravity theories for $N = 2, 3, \ldots, 8$. (Theories with $N > 8$ would contain spins not less than $\frac{5}{2}$ for which consistent gravitationally coupled Lagrangians do not exist. Thus $N = 8$ represents the maximal supergravity theory possible on present ideas.)

A supersymmetric supermultiplet in an extended theory consists of a set of SO(N) multiplets of different spins. Thus the content of a massless supermultiplet for the $N = 8$ extended theory, with maximum helicity ± 2, is as follows:

helicity	$+2$	$\pm\frac{3}{2}$	± 1	$\pm\frac{1}{2}$	0
SO(8) content	1	8	28	56	70

The total number of physical states of integer (half-integer) helicity is 128. Lagrangians for massless supermultiplets (with maximum helicity ± 2) can be written down, up to and including $N = 8$. Naturally these Lagrangians would sport just one coupling parameter (the gravitational), and are expected to be globally SO(N)-invariant. Noting that the spin-1 fields in the $N = 8$ supermultiplet (28 fields) correspond to the adjoint representation of the 'internal' global SO(8), one may raise the question: Can one add supersymmetry-preserving terms to the Lagrangian that might convert SO(N)-global to SO(N)-local, with adjoint SO(N) spin-1 fields (already contained in the super-multiplet) as Yang–Mills gauges? This would then permit us to include a second coupling parameter g of the type (and magnitude) familiar in electronuclear theory. The final unified theory would *not* be a *uni-constant* theory, but it would be a *uni-supermultiplet* theory, certainly for $N = 8$.

The answer to this question appears to be in the affirmative, though the actual construction has so far been done only up to and including $N = 5$. A characteristic of all such Lagrangians is the appearance of a cosmological term of the form $\lambda h(\phi)$ with the parameter $|\lambda| \approx |g^2\kappa^{-2}|$, and a spin-$\frac{3}{2}$ 'mass' term of the general form $g\bar{\psi}^i_\mu \sigma_{\mu\nu} \psi^i_\nu f(\phi)$. Here the ϕ are the scalar fields in the supermultiplet. Thus N-extended supergravity theories that are also locally SO(N) Yang–Mills, contain *two* parameters: the gravitational κ and g or equivalently κ and λ (the cosmological constant). As we remarked earlier the cosmological constant $\lambda \sim g^2\kappa^{-2}$ is some 66 orders of magnitude larger than the empirical cosmological upper limit would permit. Notwithstanding the attractiveness of local SO(N), I shall in the remainder of this paper set $g = 0$, and consider only pure gravitational super-Lagrangians.

In this context the most important result is the construction of the $N = 8$ super-Lagrangian by Cremmer & Julia (1979) from which Lagrangians for $N < 8$ can be derived by suitable contractions. Cremmer & Julia started by noting that the $N = 8$ supergravity supermultiplet in four dimensions has the same physical content as the $N = 1$ simple supergravity multiplet in eleven dimensions, provided that in eleven dimensions the fields introduced correspond to the elf-bein e^m_n, the spinor field $\psi_{\alpha m}$ and a three-index antisymmetric tensor $A_{[mnp]}$. The independent physical degrees of freedom on reduction to four dimensions can be checked to be 128 for bosons as well as for fermions. We have, once again, the eleven-dimensional space–time of Kaluza and Klein.

The exciting part of the Cremmer–Julia construction was the discovery of hidden (on-shell) symmetries for the equations of motion as well as (off-shell) symmetries for the Lagrangian. The on-shell symmetries were found to constitute a non-compact E_7 with 133 generators; the off-shell symmetries are SU(8), rather than the humble 'internal' SO(8) from which we started. The construction uses a scalar 56×56 matrix field V of the E_7 algebra. By writing $\partial_\mu VV^{-1}$ in the form

$$\begin{pmatrix} Q_\mu & P_\mu \\ P_\mu & Q_\mu \end{pmatrix}$$

the Q_μ piece can be considered as 63 auxiliary spin-1 objects, which occur in characteristic combinations like $(\mathrm{D}_\rho - Q_\rho)$ (for example the spin-$\frac{3}{2}$ terms read

$$\epsilon^{\mu\nu\rho\kappa}\bar{\psi}^A_\mu \gamma_\nu \gamma_5 (\mathrm{D}_\rho - Q_\rho)^B_A \psi_{\kappa B}$$

where $\psi_{\kappa B}$ is the Rarita–Schwinger SU(8) octet). Thus the 63 fields composing Q_ρ might act as Yang–Mills gauge fields for an internal SU(8), if these fields possessed a propagation character.

It is important to realize that the Q_μ are not endowed with a basic kinetic energy term in the Cremmer–Julia Lagrangian.† Cremmer & Julia made the conjecture that the Q_μ may be quantum-completed, acquiring a propagation character through quantum loops. Here Cremmer & Julia draw an analogy with the well known CP^{n-1} model in two dimensions. This model starts with a Lagrangian for scalar fields ϕ_i containing a non-propagating *auxiliary* field V_μ,

$$\mathscr{L} = -\sum_1^n |(\partial_\mu - \mathrm{i}V_\mu)\,\phi|^2.$$

† These fields present an SU(8) 'gauge' in the same sense as the antisymmetric part of the vierbein field e^α_μ does, the analogy of E_7 being with GL(4, R), and of SU(8) being with O(3, 1) in the four-dimensional gravity theory of Weyl, Sciama and Kibble. (Note that GL(4, R)/O(3, 1) has dimension $16-6 = 10$; this gives the number of components of the physical graviton field in four dimensions. Likewise the coset E_7/SU(8) with dimension 133–63 = 70 represents the 70 spin-zero physical fields in $N = 8$ extended supergravity.)

The Lagrangian exhibits a $U(1)$ symmetry:

$$\phi_i \to e^{\mathrm{i}a(x)}\phi_i,$$
$$V_\mu = \tfrac{1}{2}\mathrm{i}\Sigma\phi_i^* \overleftrightarrow{\partial_\mu}\ \phi^i \to V_\mu - \partial_\mu a.$$

There is, however, no kinetic energy term for V_μ in the Lagrangian itself. It can be shown that the field V_μ does propagate, but as a consequence of radiative effects; that it then acts both as a confining and a binding field among basic scalars of the theory; and that the spectrum of the composite states exhibits an SU(n) symmetry. For the $N = 8$ supergravity theory of Cremmer & Julia, the corresponding conjecture would be that the 63 Q_μ fields do acquire SU(8) Yang–Mills propagation similarly; that they provide electronuclear type of binding and confining forces; and that the composites arising in this theory make up an infinite-dimensional unitary representation of the non-compact E_7, whose maximal compact subgroup is SU(8).

4.4. $N = 8$ *Supergravity as a superunified theory of preons*

The conjecture that the $N = 8$ supergravity theory of Cremmer & Julia represents a preonic Lagrangian has been made by E.G.M.Z. They started by noting that the attempt to use the $N = 8$ supermultiplet, with 28 spin-1 and 56 spin-$\frac{1}{2}$ objects had come to rapid grief when one realized that $\mathrm{SO}(8) \not\supset \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. An aspect of this failure is that when we decompose SO(8) relative to $\mathrm{SU_C}(3)$ and electric charge we obtain

$$\mathbf{28} = \mathbf{8}(0) + \mathbf{1}(0) + \mathbf{3}(-\tfrac{1}{3}) + \mathbf{3}(-\tfrac{1}{3}) + \mathbf{3}(\tfrac{2}{3}) + \mathbf{\overline{3}}(\tfrac{1}{3}) + \mathbf{\overline{3}}(\tfrac{1}{3}) + \mathbf{\overline{3}}(-\tfrac{2}{3}),$$
$$\mathbf{56} = \mathbf{3}(\tfrac{2}{3}) + \mathbf{3}(-\tfrac{1}{3}) + \mathbf{3}(-\tfrac{1}{3}) + \mathbf{3}(\tfrac{2}{3}) + \mathbf{6}(\tfrac{1}{3}) + 8(0) + \mathbf{1}(-1) + \mathbf{1}_{\mathrm{L}}(0) + \mathbf{1}_{\mathrm{L}}(0).$$

Thus, the $N = 8$ supermultiplet, if identified with physical particles, might, at best, accommodate u, d, s, c (colour-triplets of quarks), a colour-sextet of quarks b, a neutral spin-$\frac{1}{2}$ octet, the electron and two neutrinos, in its spin-$\frac{1}{2}$ sector, plus coloured gluons, the photon, the Z^0, and fractionally charged superheavy gauge bosons among the spin-1 particles. There, however, are no $W^\pm$, μ, τ, ν_τ or t: these would have to emerge as composites.

Now, instead, assume that the entire $N = 8$ supermultiplet consists of preons with the exception of the SU(8) singlet, the graviton; assume that preons bind into heavy composites through the operation of forces represented by the $N = 8$ super-Lagrangian, and into 'light' composites, through the effective electronuclear type of force propagated by the composite gauges Q_μ of SU(8). Assume that this SU(8) will contain (and also spontaneously break into) the physical $\mathrm{SU}(5) \times \mathrm{SU_{family}}(3)$. We now ask: What are the 'light' preonic composites? Since the 63 composite $Q^B_{\mu A}$ are expected to be *massless* gauge particles, the other light composites should clearly belong to the supermultiplet to which these 63 particles can be assigned. One could then examine what else would be contained in the supermultiplet of which the $Q^B_{\mu A}$ are members. Does it contain, in particular, light spin-$\frac{1}{2}$ composites, identifiable with three fermion families of $\mathbf{5} + \overline{\mathbf{10}}$ of SU(5)? Is this super-multiplet unique?

E.G.M.Z. have conjectured that the supermultiplet to which the $Q^B_{\mu A}$ belong is

$$[\tfrac{3}{2}]^A, [1]^A_B, [\tfrac{1}{2}]^A_{[BC]}, [0], {}^A_{[BCD]}, \ldots, [-\tfrac{5}{2}]^A + \text{T.C.P. conjugate } [\tfrac{5}{2}]_A, [2]_{AB}, [\tfrac{3}{2}]_{A[BC]}, \ldots, [-\tfrac{3}{2}]_A.$$

This multiplet contains a variety of objects of spins greater than one. Using the preonic *ansatz* stated earlier, we shall from the start assume that all composites of spins greater than one are superheavy (Planck mass). To select the light spin-$\frac{1}{2}$ composites from among the irreducible SU(8) representations $\overline{\mathbf{504}}+\mathbf{56}+\mathbf{216}+\overline{\mathbf{8}}$ contained in this supermultiplet, E.G.M.Z. start by assuming that the 'trace parts' of the multiplet (**56** and $\overline{\mathbf{8}}$) are also superheavy. Of the remaining $\overline{\mathbf{504}}$ and **216**, they then select the maximal SU(5) anomaly-free set, such that colour and electric charge are vector-like. Using these and certain other criteria, they claim that finally within **216** and $\overline{\mathbf{504}}$ there are left just three SU(5) multiplets ($\mathbf{10}+\mathbf{5}$), which may qualify as light spin-$\frac{1}{2}$ composites and which just correspond to the three known families of quarks and leptons.

E.G.M.Z. have been criticized by Derendinger *et al.* (1981), who find no convincing reason why, for example, the 'trace' multiplets were left out of consideration, or why SU(8) should break into SU(5) × SU(3). Derendinger *et al.*, adopting different criteria, motivate a two-family set of light composites of spin $\frac{1}{2}$ emerging from a peculiar set of SU(8) multiplets $\mathbf{56}+(\overline{\mathbf{8}}+\overline{\mathbf{8}}+\overline{\mathbf{8}}+\overline{\mathbf{8}}+\overline{\mathbf{8}})$ with five $\overline{\mathbf{8}}$'s. Ellis, Gaillard and Zumino (unpublished) have attempted to show that those spin-$\frac{1}{2}$ objects contained in the **216** and $\overline{\mathbf{504}}$, which they had earlier discarded, are in fact swallowed up by higher-spin representations to give to the latter their (large) masses. (An alternative descent of SU(8) into a single-family, grand-unifying $SU(4)|_{\text{flavour}} \times SU(4)|_{\text{colour}}$ mentioned in § 2.2(*e*) may also be envisaged.) The matter rests here at present, with surely more to come in this exciting $N = 8$ supergravity preonic story.†

4.5. *Infinities in extended supergravity theories*

We saw that one of the attractive features of supersymmetric theories is the mildness of their infinities, as exemplified by the vanishing of the three-loop β for $N = 4$ extended Yang–Mills supersymmetry. What is the situation for extended supergravities?

Three types of infinities have been investigated.

(*a*) The first consists of on-shell S-matrix elements. These are one-loop *finite* for all $N \leqslant 8$, and probably also two-loop *finite*. (This assumes that duality transformations of the theory continue to hold notwithstanding quantum corrections and that there are no unexpected anomalies.) For eight or more loops, there do exist counter terms, which may signal the existence of infinities for $N = 8$. Whether such infinities are absent or not can unfortunately be decided only by a calculation. (As Kallosh has shown, counter terms at the three-loop level exist for the linear $N = 8$ theory; they may, however, disappear when the full nonlinear theory is considered.)

(*b*) Assume that for all N a Yang–Mills supersymmetric coupling of the spin-1 fields in the theory can be carried through. (As stated before, such theories have explicitly been constructed up to and including $N = 5$; the new couplings (parameter g) include a cosmological term with $\lambda = g^2/\kappa^2$. Is $\beta(g) = 0$ for such theories? Equivalently, is there no infinite renormalization of the cosmological constant? If there is not, the empirically desirable value $\lambda = 0$ is stable against renormalization.

This problem was first addressed by Christensen *et al.* (1980) and then by Curtright (1981). Their one-loop result is that $\beta = 0$ for $N = 5, \ldots, 8$. Curtright's proof has already been given when we were discussing $N = 4$ Yang–Mills extended supersymmetry. To apply his formulae, say for $N = 8$, note that

$$\beta = \frac{\hbar}{96\pi^2} g^3 \sum_S \sum_C C(S)\,(1-12S^2)\,(-1)^{2S},$$

† Is the photon a composite field? Is charge conservation spontaneously violated? Does the photon have a mass and if so, is the mass related to $R^{-1}_{\text{universe}} \approx 10^{-41}$ GeV or to the energy scale where the eleven dimensions compactify to four? Why does this compactification take place?

the summation being over the quadratic Casimirs $C(S)$ of the appropriate SO(8) multiplets as well as over the helicities (S), composing the $N = 8$ supermultiplet. The appropriate $C(S)$ are given in table 3.

Curtright finds in fact that for $N > 4$ *any* supermultiplet gives vanishing convective and magnetic contributions individually to β for all internal SO(N) (and also for any 'hidden' internal SU(N), like SU(8) of Cremmer & Julia). This means that one-loop $\beta = 0$ also, for the E.G.M.Z. composite supermultiplet.

(*c*) A third type of one-loop infinity investigated by Duff & Van Nieuwenheuisen (1980) is the Euler infinity which may arise as a renormalization of the Euler number

$$\chi = \frac{1}{32\pi^2}\int d^4x\, g^{\frac{1}{2}}(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2).$$

TABLE 3

helicity (S)	$D(S)$	$C(S)$
± 2	1	0
$\pm \frac{3}{2}$	8	1
± 1	28	6
$\pm \frac{1}{2}$	56	15
0	70	20
β	0	

This infinity is connected with the trace anomaly in supergravity theories. The result of the calculations shows that one-loop infinity is absent for all $N > 3$. The important point (for example for $N = 8$ extended supergravity) is that a naive calculation would not have given a *zero* result. One must take proper account of the Lorentz character of the scalar fields in the theory. When a descent is made from eleven dimensions to four, the 70 spin-zero fields really appear as $63\phi + 7\phi_{\mu\nu} + 1\phi_{\mu\nu\rho}$ where $\phi_{\mu\nu}$ is two-index antisymmetric and $\phi_{\mu\nu\rho}$ three-index antisymmetric. The two-index antisymmetric $\phi_{\mu\nu}$ can be shown to be equivalent to a scalar field for all purposes except for the computation of its trace anomaly; likewise $\phi_{[\mu\nu\rho]}$ is trivial except for its anomaly contribution. Once this has been taken into account, the overall $N = 8$ trace anomaly vanishes and with it the possible infinity associated with the renormalization of the Euler number. One can but marvel how supergravities manage to defeat infinities, in the examples considered. This must be connected with the essential geometry behind the supergravity theories† – a subject that we are painfully and slowly beginning to understand and one that I cannot discuss in detail in this paper.

Clearly one's first reaction to the absence of infinities in supergravity is one of rejoicing. One must remember, however, that in a conventional renormalizable theory, the structure of the infinities and the high energy behaviour of a renormalizable theory are intimately related. Even if the S-matrix in the $N = 8$ extended supergravity theory is loop-by-loop finite, it is unlikely that its *high energy* behaviour for *l-loops* would have been drastically improved from what one expects for normal gravity theory (i.e. for large E, S-matrix elements tend to $\kappa^{e+l-1}E^{l+3}$, where l is the number of loops, e the number of external lines for the graph).

If one believes that all theories, including supergravity, should exhibit Froissart boundedness for cross-sections, and this may be questioned, either a loop summation should now be made, or

† See Salam (1978), where a review is presented of the fermionic extensions of space–time (superspace) in relation to supergravity.

one must hope that the 'running' gravitational constant $\bar{\kappa}(E)$ (if this can be defined in a renormalization group sense) runs like E^{-1} for large E. In this case, S-matrix elements would indeed behave in the Froissart manner we have come to expect for normal theories (i.e. $[\bar{\kappa}(E)]^{e+l-1} \times E^{l+3} \rightarrow E^{4-e}$).

How can one use renormalization group technology for estimating the running constant $\bar{\kappa}(E)$? Is the use of such a technology even necessary? Could one devise other methods for summing successive loop contributions? The renormalization group approach to gravity theories was motivated by Julve & Tonin (1978) and by Salam & Strathdee (1978). In the language of supergravity, one may write down an (extended) conformal supergravity Lagrangian which contains g^2R^2-like terms plus a Poincaré supergravity term R/κ^2. On a power-counting basis, as is well known, such a theory is conventionally renormalizable; its failing is the presence of ghosts. These can be made arbitrarily massive by letting the coupling constant g (in the g^2R^2 like term) tend to zero *after* one has solved the renormalization group equations. One may wonder under what conditions this limit $g \rightarrow 0$ is permissible. It is interesting that one of these conditions would be $\beta(g) = 0$.

We would welcome extended conformal supergravity theories to ensure that this particular β-function vanishes. The R/κ^2-term, which acts like a mass term when added to such a theory, may still need a renormalization of κ^2. We conjecture that the renormalization group machinery may then show that $\bar{\kappa}(E) \sim E^{-1}$. I realize that there is much tortuousness and wishful thinking in this conjecture but it may still be interesting to compute one-loop corrections for an extended conformal supergravity theory to see if there is any basis for entertaining the hope that the relevant $\beta(g)$ vanishes. At the very least, the limit $g \rightarrow 0$, which can then be taken, will act as a regularizer for the physical theory.†

In conclusion, supergravity theories are attractive as field theories, and on account of their superior finiteness. The $N = 8$ supergravity is attractive in combining, in one gauged multiplet, the elementary particles and the elementary forces. Most important of all, it is attractive because it seeks the meaning of the elementary charges it uses, within the still more elementary construction of an extended space–time structure with eleven bosonic dimensions. Among these charges are included the 'fermionic charges' for which the appropriate space–time extension may be the fermionic dimensions of a superspace. There is, however, just one mass scale in the theory (M_{Planck}); the severe dynamical problem of deducing all the other masses in terms M_{Planck}, is left to the future.

Experimental outlook for particle physics

I end by considering the experimental outlook for testing the ideas that have been expressed, and one must confess that it is bleak.

There are four types of experiments currently yielding data on particle physics: (*a*) accelerator

† What, if any, are the direct experimental tests of supergravities? One such test was suggested by J. Scherk: antigravitational force of repulsion between all matter, caused by spin-1 partners (gravi-photons) of gravitons. Such a force would be short range if gravi-photons are massive. If, however, this mass is tiny, anti-gravity might manifest itself over laboratory distances. After examining records of all experiments performed to verify Newtonian laws of gravitation, and also examining the limits that could come from the known accuracy of the equivalence principle experiments, Scherk concluded that anti-gravitational effects may indeed exist with a range of 10^2–10^3 m. For details, see the tragically posthumous record of Scherk's (1980) talk given at the Europhysics Study Conference held in Erice, Sicily, in March 1980.

experiments; (*b*) cosmic-ray experiments; (*c*) non-accelerator experiments and (*d*) cosmological data. We consider the prospects for each in turn.

(*a*) *Accelerators*. Let us assume the P$\bar{\text{P}}$-collider, the Tevatron, Isabelle and LEP are available for experimentation during at least part of the decade. We shall then be well off in the TeV range of energies. Between 1990–2005, one envisages the possible installation of a P$\bar{\text{P}}$-collider in the Lep tunnel and the construction of the supertevatron. With superconducting technology these might, from an optimistic viewpoint, reach 10 TeV (centre of mass). What will happen to the subject 25 years from now?

Let us consider reaching 100 TeV, the currently accepted inverse radius of the muon, as revealed by limits on $\mu \rightarrow e + \gamma$. With present accelerator technology we shall have reached a saturation (i) in the CERN- and Fermi-laboratory sites, (ii) in available funds, and (iii) most crucially in ideas for further machine design, which, let us gratefully recall, were created for our generation by far-seeing men 25 years ago.

We desperately need, on a 25-year perspective, new ideas on accelerator design. To emphasize this point, let us remember that present designs are limited by the gradients of accelerating fields, E_{acc}. These presently attain values around 1.2 MV m^{-1} and will improve to about 5 MV m^{-1} with superconducting magnets. If a credible design with the use of lasers, for example, could be made available, E_{acc} could register values of the order of 1 GV m^{-1}. (Willis, at CERN, has considered collective ion effects, which promise field gradients of the order of 3 GV m^{-1}; while Palmer estimates 2 GV m^{-1} by using surface effects of a grating, rising to 20 GV m^{-1} if gratings were permitted to be destroyed at each pulse.)

If such designs could be realized, and one must not underestimate the difficulties (laser wavelengths are in the micrometre region), a 100 TeV accelerator need be no longer than about 30 km, perhaps even as compact as 5 km.

What I am trying to emphasize is that accelerators may become extinct as dinosaurs in 25 years, unless we take heed now and invest effort on new design.

(*b*) *Cosmic-ray experiments*. The highest possible cosmic-ray energies on Earth unfortunately do not exceed 100 TeV (centre of mass). The global cosmic-ray detection effort produces no more than 300 events per year at this energy and no more than 2000 events per year at 10 TeV (centre of mass). The number of events would increase tenfold if there was a 100 km^2 coverage with detection devices, which would certainly be worthwhile until a 100 TeV accelerator becomes available, but is no substitute for investment in new accelerators and their design.

(*c*) *Non-accelerator experiments*. These include (i) search for proton decays, (ii) search for N–$\bar{\text{N}}$ oscillations, (iii) neutrino mass and oscillation experiments involving reactors, and (iv) search (also geochemical) for neutrino-less double β-decay, and are likely to provide some of the most eagerly awaited information on the distribution of intermediate mass scales. For example, each of the proton-decay modes ($P \rightarrow e^+ + \pi^0$, $P \rightarrow e^- + \pi^+ + \pi^+$, $P \rightarrow 3\nu + \pi^+$ and $P \rightarrow 3\bar{\nu} + \pi^+$), if seen, is associated with a different mass-scale (10^{14} GeV, 10^9–10^{10} GeV, 10^5 GeV). All these modes can coexist though some of them may be rare. Thus proton decay experiments will have a long lifespan, with the vast information that they and they alone can provide. There is a good case for buying real estate under the Mont Blanc for long occupancy.

(*d*) *Cosmological data*. Notwithstanding Landau's famous admonition: 'Cosmologists are often wrong, but seldom in doubt', cosmology, while also exploring other intermediate mass scales, provides our only window on masses beyond 10^{14} GeV.

After painting this bleak picture for the experimental prospects of our subject, I must admit that I am continually being amazed at how rapidly our experimental colleagues succeed in demolishing (or sometimes demonstrating) the seemingly inaccessible and often outrageous of our theoretical speculations. This continual vigilance is the glory of all science, including our own.

References (Salam)

Baaklini, N. S. & Salam, A. 1979 *Nucl. Phys.* B **153**, 237.
Banks, T., Schwimmer, A. & Yankelowicz, S. 1980 *Physics Lett.* B **96**, 263.
Bars, I. & Yankelowicz, S. 1981 Yale University preprint YTP81-04.
Christensen, S. M., Duff, M. J., Gibbons, G. W. & Rocek, M. 1980 *Phys. Rev. Lett.* **45**, 161.
Cremmer, E. & Julia, B. 1979 *Nucl. Phys.* B **159**, 141.
Curtright, T. L. 1981 University of Florida preprint UFTP 8-23.
Curtright, T. L. & Freund, P. G. O. 1979 In *Supergravity* (ed. P. Van Nieuwenheuizen & D. Z. Freedman). Amsterdam: North Holland.
Derendinger, J. P., Ferrara, S. & Savoy, C. A. 1981 CERN Preprint TH 3052.
Deser, S. & Zumino, B. 1976 *Physics Lett.* B **62**, 335.
Duff, M. J. & Van Nieuwenheuizen, P. 1980 *Physics Lett.* B (To be published.)
Ellis, J., Gaillard, M. K., Maiani L. & Zumino, B. 1980*a* In *Proceedings of the Europhysics Conference, Erice, Sicily*. New York, London: Plenum Press. (E.G.M.Z.).
Ellis, J., Gaillard, M. K. & Zumino, B. 1980*b* *Physics Lett.* B **94**, 343.
Freedman, D., Van Nieuwenheuizen, P. & Ferrara, D. 1976 *Phys. Rev.* D **13**, 3214.
Georgi, H., Quinn, H. & Weinberg, S. 1974 *Phys. Rev. Lett.* **33**, 451.
Golfand, Yu. A. & Lichtman, E. P. 1971 *Soviet Phys. JETP Lett.* **13**, 452.
Harari, H. & Seiberg, R. 1981 *Physics Lett.* **98B**, 269.
Isham, C. J., Salam, A. & Strathdee, J. 1971 *Phys. Rev.* D **3**, 867.
Julve, J. & Tonin, M. 1978 *Nuovo Cim.* B **46**, 137.
Kaluza, T. 1921 *Sber. preuss. Akad. Wiss*, p. 966.
Klein, O. 1926 *Z. Phys.* **37**, 895.
Pati, J. C. & Salam, A. 1973*a* *Phys. Rev.* D **8**, 1240.
Pati, J. C. & Salam, A. 1973*b* *Phys. Rev. Lett.* **31**, 661.
Pati, J. C. & Salam, A. 1974 *Phys. Rev.* D **10**, 275.
Pati, J. C., Rajpoot, S. & Salam, A. 1980*a* *J. Phys.* A **13**, 2509.
Pati, J. C., Salam, A. & Strathdee, J. 1975*a* *Physics Lett.* B **59**, 265.
Pati, J. C., Salam, A. & Strathdee, J. 1975*b* *Nuovo Cim.* A **26**, 72.
Pati, J. C., Salam, A. & Strathdee, J. 1980*b* *Nucl. Phys.* (To be published); International Centre for Theoretical Physics, Trieste preprint IC/80/180.
Pati, J. C., Salam, A. & Strathdee, J. 1980*c* *Nucl. Phys.* (To be published); International Centre for Theoretical Physics, Trieste preprint IC/80/183.
Pauli, W. 1933 *Annln Phys.* **18**, 337.
Salam, A. 1978 In *Proceedings of the 19th International Conference on High Energy Physics, Tokyo*. Tokyo: Physical Society of Japan.
Salam, A. & Strathdee, J. 1974 *Nucl. Phys.* B **80**, 499.
Salam, A. & Strathdee, J. 1976 *Phys. Rev.* D **14**, 2830.
Salam, A. & Strathdee, J. 1978 *Phys. Rev.* D **18**, 4596.
Scherk, J. 1980 In *Europhysics Conference on Unification of Fundamental Particle Interactions*, p. 381. London, New York: Plenum Press.
Sohnius, M. F. & West, P. 1981 Imperial College preprint ICTP/80/81-9.
Thirring, W. 1972 *Elementary particle physics*, p. 256 (ed. P. Urban). Vienna: Springer Verlag.
't Hooft, G. 1979 *Cargese lecture* (ed. M. Levy). London: Plenum Press.
Wess, J. & Zumino, B. 1973 *Nucl. Phys.* B**70**, 39.
Witten, E. 1981 Princeton University Preprint.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF