

# GEOMETRIC ALGEBRA AND COVARIANT METHODS IN PHYSICS AND COSMOLOGY

Antony M Lewis  
*Queens' College*  
*and*  
*Astrophysics Group,*  
*Cavendish Laboratory*

A dissertation submitted for the degree of Doctor of Philosophy  
in the University of Cambridge.  
September 2000

Updated 2005 with typo corrections



# Preface

This dissertation is the result of work carried out in the Astrophysics Group of the Cavendish Laboratory, Cambridge, between October 1997 and September 2000. Except where explicit reference is made to the work of others, the work contained in this dissertation is my own, and is not the outcome of work done in collaboration. No part of this dissertation has been submitted for a degree, diploma or other qualification at this or any other university. The total length of this dissertation does not exceed sixty thousand words.

Antony Lewis  
*September, 2000*



# Acknowledgements

It is a pleasure to thank all those people who have helped me out during the last three years. I owe a great debt to Anthony Challinor and Chris Doran who provided a great deal of help and advice on both general and technical aspects of my work. I thank my supervisor Anthony Lasenby who provided much inspiration, guidance and encouragement without which most of this work would never have happened. I thank Sarah Bridle for organizing the useful lunchtime CMB discussion group from which I learnt a great deal, and the other members of the Cavendish Astrophysics Group for interesting discussions, in particular Pia Mukherjee, Carl Dolby, Will Grainger and Mike Hobson.

I gratefully acknowledge financial support from PPARC.



# Contents

<b>Preface</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Geometric Algebra</b>	<b>5</b>
2.1 Definitions and basic properties . . . . .	5
2.2 Algebra of the Plane . . . . .	8
2.3 Algebra of 3-Space . . . . .	9
2.4 Useful definitions and identities . . . . .	10
2.5 Rotations and reflections . . . . .	11
2.6 Spacetime Algebra . . . . .	12
2.7 Geometric Calculus . . . . .	14
2.8 Electromagnetism . . . . .	15
2.9 Linear functions and notation . . . . .	16
2.10 Conclusions . . . . .	19
<b>3 Electron Scattering in the Spacetime Algebra</b>	<b>21</b>
3.1 Introduction . . . . .	21
3.2 The Dirac equation . . . . .	22
3.3 Dirac theory . . . . .	23
3.4 Plane waves and basis states . . . . .	24
3.5 Feynman propagators . . . . .	26
3.6 Electron scattering . . . . .	27
3.7 Positron scattering and pair annihilation . . . . .	29
3.8 Cross-sections . . . . .	29
3.9 Coulomb scattering . . . . .	31
3.10 Compton scattering . . . . .	32
3.11 Pair annihilation . . . . .	33
3.12 Second order Coulomb scattering . . . . .	34
3.13 Spin dependence and double scattering . . . . .	36
3.14 The partial spin-sum approach . . . . .	38
3.15 Conclusions . . . . .	39

<b>4</b>	<b>Gauge Theories of Gravity</b>	<b>41</b>
4.1	Introduction . . . . .	41
4.2	Poincaré-Einstein-Cartan theory . . . . .	42
4.3	Related formulations: differential forms . . . . .	47
4.4	General Relativity . . . . .	49
4.5	Extending the gauge symmetries . . . . .	51
4.6	Weyl-Cartan gravity . . . . .	51
4.7	The field equations . . . . .	53
4.8	The Einstein gauge . . . . .	55
4.9	Discussion . . . . .	57
4.10	Conclusions . . . . .	58
<b>5</b>	<b>Quadratic Lagrangians and Topology</b>	<b>61</b>
5.1	Introduction . . . . .	61
5.2	Topological invariants . . . . .	62
5.3	Relation to instantons . . . . .	63
5.4	Topological terms for the $\bar{h}$ -field . . . . .	65
5.5	Quadratic Lagrangians . . . . .	65
5.6	Conclusions . . . . .	68
<b>6</b>	<b>Perturbations in Cosmology</b>	<b>69</b>
6.1	Introduction: Evolution of the Universe . . . . .	70
6.2	Covariant perturbations: The 3+1 decomposition . . . . .	73
6.3	General Relativity . . . . .	75
6.4	The spatial curvature . . . . .	78
6.5	The background equations . . . . .	79
6.6	Scalar, Vector, Tensor decomposition . . . . .	80
6.7	The scalar equations . . . . .	80
6.8	Evolution of the potential . . . . .	82
6.9	Perturbations in single-field inflation . . . . .	83
6.10	Quantum generation of perturbations . . . . .	85
6.11	PSTF expansion of the distribution function . . . . .	87
6.12	The Boltzmann equation . . . . .	88
6.13	Thomson scattering . . . . .	90
6.14	The scalar multipoles . . . . .	92
6.15	The fluid components . . . . .	92
6.16	The scalar harmonics . . . . .	93
6.17	The line of sight solution . . . . .	95
6.18	Instantaneous recombination approximation . . . . .	96
6.19	The CMB power spectrum . . . . .	97
6.20	Initial conditions . . . . .	99
6.21	Massive neutrinos . . . . .	100
6.22	The tensor equations . . . . .	101
6.23	Conclusion . . . . .	102



<b>7</b>	<b>Efficient computation of CMB anisotropies</b>	<b>105</b>
7.1	Introduction . . . . .	105
7.2	Modifying CMBFAST . . . . .	106
7.3	Computing the line of sight integral . . . . .	107
7.3.1	Computing the sources . . . . .	107
7.3.2	Evaluating the Bessel functions . . . . .	108
7.4	The power spectrum: integrating over wavenumber . . . . .	111
7.5	Massive neutrinos . . . . .	111
7.6	Numerical stability and accuracy . . . . .	112
7.7	Interpolating the power spectrum . . . . .	113
7.8	Results . . . . .	113
7.9	Conclusion . . . . .	115
<b>8</b>	<b>Conclusions</b>	<b>117</b>



# Chapter 1

## Introduction

It has been known for a long time that many physical phenomena can be described accurately by mathematical models. Predictions for the outcomes of experiments are generated from sets of physical laws stated in the form of equations relating various physical quantities. To derive solutions to these equations (find out what actually happens in certain cases of interest) physicists and mathematicians often use a wide variety of mathematical techniques. These often involve a great deal of abstraction, and intermediate steps in an argument often deal with quantities that have no direct physical relevance.

In this thesis we use a variety of mathematical tools to tackle problems in quantum theory, relativity and cosmology. Our choice of mathematical tool is governed by the desire to derive results in as physical a way as possible. For example we consider objects that correspond directly with physical observables wherever possible. This allows the derivations of solutions to carry some physical meaning, often leading to a better physical insight into what is happening.

Many physical quantities have a geometric nature, and it is therefore natural to describe the physics in terms of the relevant geometric quantities. The language we choose to describe and manipulate these geometric quantities is the Geometric Algebra (GA) developed by Hestenes [1, 2]. GA was developed from the work of Clifford and Grassman and provides a framework for manipulating geometric quantities in a transparent and co-ordinate independent way. In Chapter 2 we provide a general introduction to GA and establish notation. For the applications in this thesis we use the four-dimensional GA of Special Relativity - the Spacetime Algebra. This encapsulates Special Relativity in the geometric structure of the spacetime, and provides a neat language for describing relativistic physics. We also describe geometric (tensor) calculus, and show how the GA notation relates to more familiar tensor notation.

In Chapter 3 we use the Spacetime Algebra (STA) to study the Dirac equation, the equation governing the quantum behaviour of an electron, and apply it to the scattering of particles with spin. We show how the Dirac equation arises fairly naturally as the quantized version of an equation for a classical spinning particle. Using the STA approach the spinors and observables of Dirac theory have a direct physical interpretation, and there is no need to introduce abstract Dirac matrices as in traditional approaches.

After establishing the STA form of various standard results in Dirac theory we move on to study the scattering of particles with spin. Hestenes [3] has shown how scattering processes can be described in the STA by a scattering operator that rotates and dilates initial states into final states. The analysis was extended by Lewis [4] to handle spin-dependent calculations, and we clarify and develop this work further here. We show how spin-independent results can be derived in a manifestly spin-independent way, and how spin-dependent results can be given in terms of the relevant rest spin planes. We re-derive various well-known results for the scattering of electrons from positrons, photons and Coulomb fields, handling the spin-dependence directly where necessary. We briefly show how the method could be extended to more complicated cases of spin-dependent multi-particle scattering.

In the remainder of the thesis we consider topics in gravitational physics. We make use of GA where appropriate to study geometric quantities, in particular in the STA formulation of theories of gravitation. After discussing the fundamental theory we then consider applications to perturbations in cosmology.

In Chapters 4 and 5 we discuss the formulation of gravity as a gauge theory using the STA, following the approach of Lasenby *et al.* [5]. In this approach gravity involves gauge fields defined over a flat background space, giving a theory much closer in spirit to the gauge theories of particle physics than to the curved-spacetime approach of General Relativity (GR). However the spacetime that we observe still appears curved — observable quantities will involve non-trivial gauge fields and can be consistent with the predictions of GR. In the gauge theory approach one assumes the existence of local symmetries that determine the nature of the gauge fields that must be present. The dynamics of the gauge fields are determined from an action principle — the minimization of an action gives rise to a set of gravitational field equations that describe how the fields behave. To construct a theory one has to decide on which symmetries to gauge, and then which action to use. The symmetries constrain the possible form the action could take, but in general there are a variety of possibilities.

In Chapter 4 we show that by imposing the minimum number of symmetries and using the simplest action one obtains Einstein-Cartan theory — a theory that is locally equivalent to GR in the absence of torsion. We review some of the key results of GR in the STA formulation and show how our notation relates to that of differential forms. We then consider the effect of including an additional scaling symmetry and discuss the corresponding action. We show how a certain choice of action leads to a theory that is equivalent to GR with the addition of a massive vector field and cosmological constant. We discuss how our gauge theory formulation relates to similar scale invariant theories in other theories of gravity.

Even after introducing an additional scaling symmetry the gravitational action is not well constrained by the symmetries and there are still a wide variety of possibilities. We discuss some of these in Chapter 5, systematically constructing the various possible independent quadratic terms for the action. These terms form legitimate contributions to the gravitational action in both scale invariant and non-scale invariant theories, and could potentially be significant at high energies.

We also exploit the construction of gravity as a gauge theory to consider analogues of

the topological structures encountered in Yang-Mills gauge theory. Instanton numbers for windings of one of the gravitational gauge fields provide the gauge theory analogue of the topological invariants of GR. The topological invariants are described by quadratic action integrals and do not contribute to the local field equations. They therefore also constrain the number of possible quadratic terms in the classical gravitational action, and we derive the field equations obtained from the five remaining independent parity non-violating terms.

In gravitational theory one of the gauge symmetries is a local displacement symmetry, so the dynamical equations governing the physics must be invariant under  $x \rightarrow x'$ . Gauge invariant equations will be made up of covariant quantities, quantities that transform as  $F(x) \rightarrow F(x')$  under the displacement and also transform homogeneously under the other gauge symmetries present. It is these variables that we need to construct observables, and covariant quantities are therefore our preferred variables as they have some direct physical meaning.

In Chapter 6 we assume that GR accurately describes cosmology and restrict our attention to a gauge theory of gravity that is equivalent to GR. The universe appears to be approximately homogeneous and isotropic on large scales and we can describe the structure in the universe in terms of perturbations about an exactly homogeneous and isotropic model. We employ covariant variables to quantify these perturbations and discuss their evolution, avoiding problems with gauge ambiguities that can arise using other methods. We use the notation established in the earlier chapters to review previous work on covariant perturbation theory [6–10], deriving a set of physically transparent linearized equations. These can be split into equations for scalar and tensor modes as required. We review how to perform a harmonic expansion of the scalar equations and compute the Cosmic Microwave Background (CMB) power spectrum.

It is thought that cosmological perturbations originated from quantum fluctuations during a period of cosmic inflation. We derive covariant equations for the propagation of perturbations during single-field inflation and show how the covariant variables relate to the gauge invariant variables used in other approaches [11]. We describe how the perturbations present during inflation can be related to those present after inflation, and review the covariant multipole analysis of the Boltzmann equation needed to study their subsequent propagation. We follow the approach of Ellis *et al.* [10] to derive an infinite two-dimensional hierarchy of energy-integrated multipole equations, and show how these can be applied to the various species of matter present. A Green's function solution to the photon hierarchy can be used to calculate the perturbations today in terms of a line-of-sight integral.

We give a new covariant analysis of massive neutrino perturbations, deriving equations for the propagation of the distribution function and of the energy-integrated multipoles. Non-covariant equations for the distribution function are well known [12], however making judicious use of the energy integrated equations we provide a much more efficient scheme for propagating the equations accurately. Similar schemes could be used for the propagation for other forms of non-relativistic matter.

Performing a harmonic expansion of the covariant perturbation equations we arrive at a set of equations that can be implemented numerically. In Chapter 7 we describe a numerical code for computing CMB anisotropies using of the results of Chapter 6. We modify the popular CMBFAST code [13, 14], using the line-of-sight integral to compute accurate results

efficiently. As well as implementing the covariant equations we also extend the code to handle closed models (those with compact spatial sections), allowing us to compute new predictions for the CMB power spectra in these models. We also implement the covariant massive neutrino equations, using a truncated expansion in velocity dispersion to propagate the perturbations efficiently. The code we develop provides a semi-independent check on the results of CMBFAST as well as being significantly faster in many cases. The code was made publicly available and has already proved useful for constraining the curvature of the universe from new observational data [15–18].

Throughout we use the summation convention and ‘natural’ units where  $c = \epsilon_0 = \hbar = 1$ .

## Chapter 2

# Geometric Algebra

Geometric Algebra (GA) forms a powerful mathematical language for studying geometric objects. In physics we are often concerned with such things, for example a velocity can be represented by a vector. Similarly a rotation can be represented by the plane of rotation and a magnitude of rotation. This latter object is called a bivector, the geometric quantity describing plane-like objects without reference to an arbitrary co-ordinate system. GA provides a set of tools for manipulating such objects so physical problems can frequently be studied directly in terms of objects that most naturally describe the physics.

Mathematically GA is a type of Clifford Algebra, invented (or discovered) in its mathematical guise in the 19th century. David Hestenes was the first major proponent of GA, his book *Spacetime Algebra* [2] containing many applications in relativistic physics. A second book, *Clifford Algebra to Geometric Calculus* [1], has a rigorous and fairly complete treatment of the maths used in many subsequent applications. In this thesis we make extensive use of GA where appropriate. However we develop a notation that should allow readers unfamiliar with GA to map most of our results more or less directly into tensor or differential forms notation if they prefer.

This chapter constitutes a fairly thorough introduction to the foundations of the algebra, and we study some concrete examples gradually building up from the algebra of the plane to the full spacetime algebra. We develop geometric calculus and a notation for linear functions, and show how these relate to standard tensor notation.

### 2.1 Definitions and basic properties

To understand Geometric algebra we shall need a few new concepts. We are familiar with scalars, and with vectors, directed line segments. We now introduce a bivector as a directed *plane* segment. This is not a scalar or a vector but a new object. In fact we can introduce an entire set of  $r$ -vectors, including directed volumes and directed hypervolumes. We shall see later how these arise.

GA gives a way to manipulate *multivectors*. What is a multivector? It is a collection (or sum) of  $r$ -vector parts, so it will have a scalar part, a vector part, a bivector part, and so on. A multivector is a convenient way of dealing with the different component types

all in one go. The different  $r$ -vector components of a multivector are defined to be *graded* corresponding to their dimensionality: scalars are grade 0, vectors grade 1, bivectors grade 2, *etc.* We need some notation to represent the grade  $r$  component (the  $r$ -vector part) of a multivector  $A$ , which we write as  $\langle A \rangle_r$ . The scalar part of  $A$  is written for simplicity just as  $\langle A \rangle = \langle A \rangle_0$ . We can therefore write

$$A = \langle A \rangle + \langle A \rangle_1 + \langle A \rangle_2 + \dots \quad (2.1)$$

The separate parts of a multivector are maintained separately during addition, as is done for the real and imaginary parts of complex numbers — scalar parts only add to scalar parts, vector parts only to vector parts, *etc.* We shall assume that the algebra of the scalars is equivalent to that of the real numbers — complex numbers are not needed [1, 19]. It should be stressed that vectors and scalars in geometric algebra are exactly the same as the ones we are used to. GA just gives new ways of combining and manipulating them.

It is the definition of multiplication which makes the algebra interesting. Define the *geometric product* of multivectors  $A$ ,  $B$ , and  $C$  by the following rules

1.  $A(BC) = (AB)C$  (associative)
2.  $A(B + C) = AB + AC$  (distributive)

Though the geometric product of multivectors is not commutative scalars still commute with everything. There is also the additional rule for vectors:

3. For a vector  $a$ , the square  $a^2$  is equal to  $a \cdot a$ .

Since  $a \cdot a$  is just the usual scalar product this third rule is important as it relates the algebra to measurable quantities. Using it we can derive a general formula for the scalar product of two vectors. Since  $a + b$  is a vector if  $a$  and  $b$  are vectors, the third rule gives

$$(a + b)^2 = (a + b) \cdot (a + b). \quad (2.2)$$

Expanding both sides we get

$$a^2 + b^2 + ab + ba = a^2 + b^2 + 2a \cdot b \Rightarrow (ab + ba) = 2a \cdot b. \quad (2.3)$$

So in geometric algebra the scalar *inner product* of two vectors is given by

$$a \cdot b = \frac{1}{2}(ab + ba). \quad (2.4)$$

The inner product is symmetric, so we define an antisymmetric product

$$a \wedge b \equiv \frac{1}{2}(ab - ba) \quad (2.5)$$

so that

$$ab = a \cdot b + a \wedge b \quad (2.6)$$



The antisymmetric product  $a \wedge b$  is called the *outer product*. To see what sort of multivector  $a \wedge b$  is we can calculate its square

$$(a \wedge b)^2 = \frac{1}{4}(ab - ba)(ab - ba) = \frac{1}{4}((ab + ba)^2 - 4a^2b^2) = -a^2b^2 \sin^2 \theta \quad (2.7)$$

using the previous result that  $ab + ba = 2a \cdot b = 2|a||b| \cos \theta$ . This square is negative, so since scalars are supposed to be real  $a \wedge b$  cannot be a scalar. Also, since rule three tells us that the square of a vector should be the square of its length, it cannot be a vector either.<sup>1</sup> Since two vectors define a plane we define  $a \wedge b$  to be a grade 2 *bivector*, a directed area in the plane of  $a$  and  $b$ . The magnitude of the area is the same as the magnitude of the cross product in three dimensions. However the result of the outer product is *not* a vector — it is a directed area. The outer product is defined in any dimension greater than one, making it much more powerful than the cross product which is only defined in 3D.

A bivector  $a \wedge b$  can be interpreted as the area swept out when  $a$  is moved along  $b$ , in the plane of the two vectors. Alternatively one can view the area simply as a circle in the plane of the vectors, in which there is a clockwise or anti-clockwise sense. The product  $a \wedge b \wedge c$  (defined below) can be visualized as the volume swept out when the area  $a \wedge b$  is swept out along  $c$ . This helps to see that  $a \wedge b \wedge c$  is zero if the vectors are not linearly independent. A bivector of the form  $a \wedge b$  is called a *blade*. A general bivector can be written as the sum of various blades; only in two or three dimensions can all bivectors can be written as a single blade.

We have shown that  $ab = \langle ab \rangle$  is a scalar and that  $a \wedge b = \langle ab \rangle_2$  is a bivector, so the inner product by a vector is a grade lowering operator, and the outer product by a vector is a grade raising operator. We can extend this property to define the outer and inner products of general multivectors to be grade raising and grade lowering. For example we want  $a \wedge b \wedge c$  mentioned above to be a grade three volume element. Writing a grade  $r$  multivector  $A$  as  $A_r$  we therefore define

$$A_r \wedge B_s \equiv \langle AB \rangle_{r+s} \quad (2.8)$$

$$A_r \cdot B_s \equiv \langle AB \rangle_{|r-s|}. \quad (2.9)$$

Note that the relation  $ab = a \cdot b + a \wedge b$  does not hold when  $a$  and  $b$  become general multivectors. However it can be shown [1] that if one of the multivectors is a vector  $a$

$$a \cdot A_r \equiv \langle a A_r \rangle_{r-1} = \frac{1}{2}(a A_r - (-1)^r A_r a) \quad (2.10)$$

$$a \wedge A_r \equiv \langle a A_r \rangle_{r+1} = \frac{1}{2}(a A_r + (-1)^r A_r a) \quad (2.11)$$

so we have  $a A_r = a \cdot A_r + a \wedge A_r$ . We see that the symmetry of the inner and outer products with a vector alternates with increasing grade. The outer product is associative so  $a \wedge (b \wedge c) = (a \wedge b) \wedge c = a \wedge b \wedge c$  as can readily be checked. The inner product for arbitrary multivectors can be defined by taking the inner products of the components:

<sup>1</sup>This argument of course only applies in a space with positive definite signature. However it provides motivation for *defining*  $a \wedge b$  to be a bivector in any space.

$A \cdot B = \sum_{r,s} \langle A \rangle_r \cdot \langle B \rangle_s$ , and similarly for the outer product. The inner product with a scalar is taken to be zero; the outer product with a scalar to be the product with the scalar.

Each part of a multivector can be split up into a sum of linearly independent components of the same type — just as one can express a vector as a sum over basis vectors, one can express bivectors as a sum over basis bivectors. One usually chooses a set of orthonormal basis vectors to describe the space. These can then be multiplied together to build up the basis elements of the full geometric algebra. To see what the algebra actually looks like in a concrete case we now examine the algebra of the plane.

## 2.2 Algebra of the Plane

In 2D there are two orthonormal basis vectors  $\sigma_1$  and  $\sigma_2$ , where  $\sigma_1 \cdot \sigma_2 = 0$ . The algebra has one scalar element 1, two vector elements  $\sigma_1$  and  $\sigma_2$ , and one bivector element  $\sigma_1 \sigma_2 = \sigma_1 \wedge \sigma_2$ . We cannot generate any further independent quantities: since the basis vectors are orthogonal they anticommute and  $\sigma_1^2 = \sigma_2^2 = 1$ , so

$$\sigma_1 \sigma_2 \sigma_1 = -\sigma_1 \sigma_1 \sigma_2 = -\sigma_2 \quad (2.12)$$

$$\sigma_1 \sigma_2 \sigma_2 = \sigma_1 \quad (2.13)$$

$$(\sigma_1 \sigma_2)^2 = \sigma_1 \sigma_2 \sigma_1 \sigma_2 = -\sigma_1 \sigma_2 \sigma_2 \sigma_1 = -1. \quad (2.14)$$

All higher products of the basis vectors reduce to one of our four basis elements.

A general multivector will be a sum of these basis elements:

$$A = \alpha + \beta \sigma_1 + \gamma \sigma_2 + \delta \sigma_1 \sigma_2 \quad (2.15)$$

where the Greek letters are scalars. Multiplication of multivectors takes place element by element, so

$$AB = (\alpha_a + \beta_a \sigma_1 + \gamma_a \sigma_2 + \delta_a \sigma_1 \sigma_2)(\alpha_b + \beta_b \sigma_1 + \gamma_b \sigma_2 + \delta_b \sigma_1 \sigma_2). \quad (2.16)$$

This gives, for example, for the scalar part of the product

$$\langle AB \rangle = \alpha_a \alpha_b + \beta_a \beta_b + \gamma_a \gamma_b - \delta_a \delta_b. \quad (2.17)$$

The other parts of the product are generated similarly. Note that the bivector  $\sigma_1 \sigma_2$  has a negative square. Indeed, the even grade elements of the algebra  $z = x + y \sigma_1 \sigma_2$  form a subalgebra equivalent to the complex numbers:

$$z_1 z_2 = (x_1 + y_1 \sigma_1 \sigma_2)(x_2 + y_2 \sigma_1 \sigma_2) = (x_1 x_2 - y_1 y_2) + (x_1 y_2 + y_1 x_2) \sigma_1 \sigma_2. \quad (2.18)$$

This shows that anything that can be done using complex numbers can be done equally well using the geometric algebra of the plane. Sometimes this can be used to give a direct geometric interpretation to equations which conventionally contain the scalar imaginary  $i$ .

The outer product of two vectors  $a$  and  $b$  can be expressed as  $a \wedge b_\perp$ , where  $b_\perp$  is the projection of  $b$  perpendicular to  $a$  (the outer product with the parallel component

is zero). The component of  $b$  perpendicular to  $a$  has magnitude  $|b| \sin \theta$ , showing that  $a \wedge b = \sigma_1 \sigma_2 |a| |b| \sin \theta$ , and therefore that  $|a \wedge b| = |a| |b| \sin \theta$ . For any unit vectors  $m$  and  $n$

$$mn = m \cdot n + m \wedge n = \cos \theta + m \wedge n_{\perp} = \cos \theta + \sigma_1 \sigma_2 \sin \theta. \quad (2.19)$$

Because of the isomorphism with complex numbers we can write

$$mn = \cos \theta + \sigma_1 \sigma_2 \sin \theta = e^{\sigma_1 \sigma_2 \theta} = e^B, \quad (2.20)$$

where  $B$  is a bivector with magnitude  $\theta$ .

## 2.3 Algebra of 3-Space

We now have an additional orthonormal basis vector  $\sigma_3$ , and we can generate the following basis elements:

$$\begin{array}{cccc} 1 & \{\sigma_1, \sigma_2, \sigma_3\} & \{\sigma_1 \sigma_2, \sigma_2 \sigma_3, \sigma_3 \sigma_1\} & \sigma_1 \sigma_2 \sigma_3 \\ \text{scalar} & \text{vectors} & \text{bivectors} & \text{trivector} \end{array}$$

It is easy to show that the trivector element commutes with all vectors, and hence that it commutes with all multivectors. Also  $(\sigma_1 \sigma_2 \sigma_3)^2 = \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 \sigma_3 = -1$  by anti-commuting the products. The trivector is also the highest element of the space. Highest grade elements are called *pseudoscalars*, and, since there are no higher grades, their outer products with vectors are zero. We give this pseudoscalar the symbol

$$I \equiv \sigma_1 \sigma_2 \sigma_3. \quad (2.21)$$

The algebra of  $z = x + Iy$ , where  $x$  and  $y$  are scalars, is equivalent to the complex numbers. In fact we are spoilt for choice for  $\sqrt{-1}$  since any unit bivector also squares to give  $-1$ .

Using the definition of  $I$  we get that

$$I\sigma_1 = \sigma_2 \sigma_3, \quad I\sigma_2 = \sigma_3 \sigma_1, \quad \text{and} \quad I\sigma_3 = \sigma_1 \sigma_2. \quad (2.22)$$

So we see that multiplication by the pseudoscalar defines the *duality* operation - it maps from an  $r$ -grade element to an  $(n - r)$  grade element, where  $n$  is the number of dimensions. In 3D the duality operation maps from bivectors to vectors (and vice-versa). For this reason it is sometimes useful to write the bivectors as  $\Sigma_k \equiv I\sigma_k$ .

Using the duality operation in 3D we can recover the traditional cross product from the outer product:

$$\mathbf{a} \times \mathbf{b} = -I\mathbf{a} \wedge \mathbf{b}. \quad (2.23)$$

For example  $\sigma_1 \times \sigma_2 = -I\sigma_1 \sigma_2 = -II\sigma_3 = \sigma_3$  as we expect. Here we adopt the convention that inner and outer products take place before geometric products, so  $I\mathbf{a} \wedge \mathbf{b} = I(\mathbf{a} \wedge \mathbf{b})$ .

## 2.4 Useful definitions and identities

Geometric algebra gives us a way to define inverses of vectors. Since  $a^2$  is a scalar, as long as it is non-zero we can define

$$a^{-1} = \frac{a}{a^2} \quad (2.24)$$

so that  $a^{-1}a = 1$ .

We define the *reverse* of a multivector  $A$ , written  $\tilde{A}$ , which reverses all the vector products making up the multivector so that

$$(AB)\tilde{\phantom{AB}} = \tilde{B}\tilde{A}, \quad (2.25)$$

$$\tilde{a}_i = a_i, \quad (2.26)$$

$$(a_1 a_2 \dots a_r)\tilde{\phantom{a_1 a_2 \dots a_r}} = a_r \dots a_2 a_1 \quad (2.27)$$

This is used to define the modulus of multivectors:

$$|A| = \langle \tilde{A}A \rangle^{\frac{1}{2}}. \quad (2.28)$$

which correctly gives  $|a \wedge b| = |a||b| \sin \theta$  from (2.7).

As an example of using the algebra to derive useful identities let us look at the GA equivalent of the vector triple product,  $a \cdot (b \wedge c)$ . From  $ab = a \cdot b + a \wedge b$  we know that

$$ab = -ba + 2a \cdot b \quad (2.29)$$

so

$$a \cdot (b \wedge c) = \langle ab \wedge c \rangle_1 = \frac{1}{2} \langle abc - acb \rangle_1 = \frac{1}{2} \langle -bac + 2a \cdot bc + cab - 2a \cdot cb \rangle_1 \quad (2.30)$$

$$= a \cdot bc - a \cdot cb. \quad (2.31)$$

Here we have used the fact that  $\langle cab \rangle_1 = \langle bac \rangle_1$  because vectors reverse to themselves. Specializing to 3D we can convert this to the traditional relation for  $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$  as follows:

$$\mathbf{a} \cdot (\mathbf{b} \wedge \mathbf{c}) = \langle \mathbf{ab} \wedge \mathbf{c} \rangle_1 = \langle \mathbf{a}(\mathbf{Ib} \times \mathbf{c}) \rangle_1 = \langle \mathbf{Iab} \times \mathbf{c} \rangle_1 = \mathbf{Ia} \wedge (\mathbf{b} \times \mathbf{c}) \quad (2.32)$$

$$= -\mathbf{a} \times (\mathbf{b} \times \mathbf{c}). \quad (2.33)$$

The GA identity is however quite general. We have also managed to prove the identity entirely algebraically without resorting to components.

Identities for more general multivectors take more work. We list some of the most useful here; proofs can be found in [1]

$$a \cdot (A_r \wedge B_s) = (a \cdot A_r) \wedge B_s + (-1)^r A_r \wedge (a \cdot B_s) \quad (2.34)$$

$$a \wedge (A_r \cdot B_s) = (a \cdot A_r) \cdot B_s + (-1)^r A_r \cdot (a \wedge B_s) \quad (2.35)$$

$$A_r \cdot (B_s \cdot C_t) = (A_r \wedge B_s) \cdot C_t \quad r + s \leq t, \quad r, s > 0. \quad (2.36)$$

Finally, two more items of notation are useful. The scalar product of multivectors we write as

$$A*B \equiv \langle AB \rangle \quad (2.37)$$

which is sometimes more convenient than using angle brackets. The commutator product is written

$$A \times B \equiv \frac{1}{2}(AB - BA) \quad (2.38)$$

(not to be confused with the traditional cross product). Clearly for vectors this coincides with the outer product. If  $A$  is a vector and  $B$  is a bivector then it is the same as the inner product. However for general multivectors the commutator product will not coincide with either the inner or outer product.

## 2.5 Rotations and reflections

Geometric algebra provides a neat way of handling reflections and rotations in any number of dimensions. Given a unit vector  $n$ , any vector can be resolved into parts parallel and perpendicular to it, so that  $a = a_{\parallel} + a_{\perp}$ . Since  $a_{\perp}$  anticommutes with  $n$  it follows that the reflection in the hyperplane orthogonal to  $n$  is given by

$$a_{\perp} - a_{\parallel} = (a_{\perp} - a_{\parallel})nn^{-1} = -nan^{-1}. \quad (2.39)$$

If we have a second unit vector  $m$  where  $m^2 = n^2$  we can make a rotation by combining the two reflections

$$a \rightarrow mnan^{-1}m^{-1} = mnanm. \quad (2.40)$$

We define a *rotor*  $R \equiv mn$ , so that

$$a \rightarrow Ra\tilde{R}. \quad (2.41)$$

Here  $R$  is a *simple rotor*, one that can be written as the product of two vectors. In a general space where not all unit vectors satisfy  $n^2 = m^2$  (as in the Spacetime Algebra) a general rotation can be described by a rotor made up of a product of simple rotors. In all cases the same transformation law applies,  $R$  contains only even elements, and  $R\tilde{R} = \tilde{R}R = 1$ . As a check the scalar product is invariant under rotations

$$(Ra\tilde{R}) \cdot (Rb\tilde{R}) = \frac{1}{2}(Ra\tilde{R}Rb\tilde{R} + Rb\tilde{R}Ra\tilde{R}) = R(a \cdot b)\tilde{R} = a \cdot b \quad (2.42)$$

since  $a \cdot b$  is a scalar and commutes with  $R$ .

As we saw earlier in (2.20), the product of two Euclidean unit vectors can be expressed as  $e^B$ , where  $B$  is a bivector in the plane of rotation. The magnitude of  $B$  was the angle  $\theta$  between  $n$  and  $m$ , which is in fact half the rotation angle. So we can write

$$R = e^{-B/2} \quad (2.43)$$

where  $B$  is a bivector in the plane of the rotation, and has magnitude equal to the rotation angle.

Note that the rotor  $R$  gives the rotation in terms of the plane in which it takes place. This is more general than the procedure used in three dimensions of defining a rotation by a vector perpendicular to the plane of the rotation — in higher dimensions the perpendicular vector is ambiguous, whereas the plane of rotation is not. The rotation also works with any vector product (and hence with any multivector), whatever grade, since

$$a_1 a_2 \dots a_r \rightarrow R a_1 \tilde{R} R a_2 \tilde{R} \dots R a_r \tilde{R} = R a_1 a_2 \dots a_r \tilde{R}. \quad (2.44)$$

In 3-space we can map the bivector to an orthogonal vector by the duality operation, giving

$$R = e^{-Ia/2} = \cos(|a|/2) - I\hat{a} \sin(|a|/2) \quad (2.45)$$

for a rotation of  $|a|$  radians about the  $a$  axis, where  $\hat{a}$  is a unit vector in the  $a$  direction. The vector part of the sine term can be pulled out of the sine, since the sine series has only odd powers and each term in the series is a vector. Since  $R = e^{-B/2}$ , and  $B$  is a bivector (so  $\tilde{B} = -B$ ), the reverse is  $\tilde{R} = e^{B/2}$ . This gives the simple rotation of a multivector  $A$ , by  $|a|$  radians about the  $a$  axis, as

$$A \rightarrow e^{-Ia/2} A e^{Ia/2}. \quad (2.46)$$

## 2.6 Spacetime Algebra

Minkowski Spacetime is four dimensional so to discuss relativity we need to develop a four dimensional geometric algebra, the Spacetime Algebra (STA). In the STA the squares of the basis vectors are no longer all positive: we shall use the metric  $(+ - - -)$ , with the four basis vectors  $\gamma_\mu$ , where  $\gamma_0^2 = 1$ , and  $\gamma_k^2 = -1$  for  $k = 1, 2, 3$ . The basis vectors then satisfy the Dirac algebra

$$\gamma_\mu \cdot \gamma_\nu \equiv \frac{1}{2}(\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) = \text{diag}(+ - - -). \quad (2.47)$$

By repeated multiplication by the basis vectors we get the 16 elements of the STA:

$$\begin{array}{cccccc} 1 & \{\gamma_\mu\} & \{\sigma_k, \Sigma_k\} & \{I\gamma_\mu\} & I \\ \text{scalar} & \text{vectors} & \text{bivectors} & \text{pseudovectors} & \text{pseudoscalar} \end{array}$$

The bivectors  $\sigma_k \equiv \gamma_k \gamma_0$  are isomorphic to the basis vectors we used for 3-space. For example

$$I\sigma_3 = \sigma_1 \sigma_2 \sigma_3 \sigma_3 = \gamma_1 \gamma_0 \gamma_2 \gamma_0 \gamma_3 \gamma_0 \gamma_3 \gamma_0 = \gamma_1 \gamma_0 \gamma_2 \gamma_0 = \sigma_1 \sigma_2. \quad (2.48)$$

Similarly the  $\Sigma_k = I\sigma_k$  are isomorphic to the basis bivectors of Euclidean 3-space. In the STA we have the definition

$$I \equiv \sigma_1 \sigma_2 \sigma_3 = \gamma_1 \gamma_0 \gamma_2 \gamma_0 \gamma_3 \gamma_0 = \gamma_0 \gamma_1 \gamma_2 \gamma_3 \quad (2.49)$$

consistent with our earlier definition of the pseudoscalar as the highest grade element. Note that though  $I$  commutes with  $\sigma_k$ , it anticommutes with vectors and trivectors. The isomorphism shows that anything that can be done in the 3D algebra can equally well be done using the even elements of the STA. If we take  $\gamma_0$  to be the lab frame velocity vector

the  $\sigma_k$  then represent a frame in space relative to the  $\gamma_0$  vector. That vectors in three space should be represented by bivectors is not unreasonable since vectors in 3D are line segments existing for a certain time, which gives a bivector in the STA.

We can do a space-time split of a vector into the  $\gamma_0$  frame by multiplying by  $\gamma_0$ . Bold letters are now used for relative 3-vectors (spacetime bivectors). The split is

$$a\gamma_0 = a_0 + \mathbf{a} \quad (2.50)$$

where  $a_0 = a \cdot \gamma_0$  and  $\mathbf{a} = a \wedge \gamma_0$ . In particular  $x$  is split (using natural units)

$$x\gamma_0 = x \cdot \gamma_0 + x \wedge \gamma_0 = t + \mathbf{x}. \quad (2.51)$$

If we define<sup>2</sup>  $\beta = v \cdot \gamma_0 = dt/d\tau$ , then the relative 3-velocity is

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = \frac{d\tau}{dt} \frac{d\mathbf{x}}{d\tau} = \frac{v \wedge \gamma_0}{\beta}. \quad (2.52)$$

Using this we have that

$$v\gamma_0 = v \cdot \gamma_0 + v \wedge \gamma_0 = \beta(1 + \mathbf{v}). \quad (2.53)$$

Since  $v^2 = 1$  is invariant we get

$$1 = (v\gamma_0)(\gamma_0 v) = \beta(1 + \mathbf{v})\beta(1 - \mathbf{v}) = \beta^2(1 - \mathbf{v}^2) \quad (2.54)$$

so rearranging we get the familiar result

$$\beta = \frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \mathbf{v}^2}}. \quad (2.55)$$

Using the space-time split of  $x$  we have

$$x^2 = x\gamma_0\gamma_0 x = (t + \mathbf{x})(t - \mathbf{x}) = t^2 - \mathbf{x}^2. \quad (2.56)$$

General Lorentz transformations can be expressed as four-dimensional rotations, expressed in GA by the rotor transformation law. In four dimensions we cannot do all possible transforms by just one rotation in a plane, so we split up the rotor into two simple rotors — a spatial rotation followed by a *Lorentz boost*. The boost is just a rotation in a time-like plane, used to change the relative velocity. A boost  $L$  will rotate the vector  $\gamma_0$  to the new velocity  $v$ ,

$$v = L\gamma_0\tilde{L}. \quad (2.57)$$

We can use (2.40) to make the rotor out of two reflections: reflect in the hyperplane orthogonal to the line between the two vectors  $(\gamma_0 + v)/|\gamma_0 + v|$ , and then reflect in the hyperplane orthogonal to  $v$ . This gives

$$L = v \frac{\gamma_0 + v}{|\gamma_0 + v|} = \frac{1 + v\gamma_0}{\sqrt{2(1 + v \cdot \gamma_0)}}. \quad (2.58)$$

---

<sup>2</sup>Though it is more usual to use the symbol  $\gamma$  we use  $\beta$  here to avoid confusion with the basis vectors

Squaring this we see that  $L^2 = v\gamma_0$ , which, from (2.53), gives  $L^2 = \beta(1 + \mathbf{v})$ . Using the space-time splits of the velocity,  $L$  becomes

$$L = \frac{1 + \beta + \beta\mathbf{v}}{\sqrt{2(1 + \beta)}}. \quad (2.59)$$

Let us see what the parts of a vector look like after a boost. Using  $\tilde{L}\gamma_0 = \gamma_0 L$  we get

$$a'_0 + \mathbf{a}' = La\tilde{L}\gamma_0 = La\gamma_0 L = L(a_0 + \mathbf{a})L. \quad (2.60)$$

If we now split  $\mathbf{a}$  into parts parallel and orthogonal to  $\mathbf{v}$  (the only non-commuting part of  $L$ ), we have

$$a'_0 + \mathbf{a}' = L^2(a_0 + \mathbf{a}_{\parallel}) + L\tilde{L}\mathbf{a}_{\perp} \quad (2.61)$$

$$= \beta(1 + \mathbf{v})(a_0 + \mathbf{a}_{\parallel}) + \mathbf{a}_{\perp} \quad (2.62)$$

Equating parts of the same grade we recover the Lorentz transformations

$$a'_0 = \beta(a_0 + \mathbf{v} \cdot \mathbf{a}) \quad (2.63)$$

$$\mathbf{a}' = \beta(\mathbf{a}_{\parallel} + a_0\mathbf{v}) + \mathbf{a}_{\perp}. \quad (2.64)$$

## 2.7 Geometric Calculus

Geometric Algebra provides a framework for extending calculus to provide many useful generalizations of standard results and can provide an alternative to traditional tensor calculus. The subject is developed extensively in [1], but we shall only look at the most important points here.

We can define a partial derivative of an arbitrary multivector valued function  $F(a)$  of a vector  $a$  in the direction of a vector  $b$ . Formally this is done by the definition:

$$b \cdot \partial_a F(a) \equiv \lim_{\tau \rightarrow 0} \frac{F(a + \tau b) - F(a)}{\tau} \quad (2.65)$$

where  $b \cdot \partial_a$  is a single scalar operator. This can be used to define the full derivative by reference to a frame  $e_i$  of vectors. A reciprocal frame,  $e^j$ , is defined such that  $e_i \cdot e^j = \delta_{ij}$  (note that  $e^j$  is also just a vector, it is not a 1-form or anything else). We now define (using the summation convention)

$$\partial_a \equiv e^i e_i \cdot \partial_a. \quad (2.66)$$

In the case where  $a$  is the position vector  $x$  we get the vector derivative

$$\nabla \equiv \partial_x = e^i e_i \cdot \partial_x. \quad (2.67)$$

If  $\nabla$  is acting on a vector then  $\nabla = e^i \partial_i$ , where  $\partial_i$  is the differential with respect to the  $e_i$  component. In two and three dimensions the reciprocal frame is the same as the frame,



but we shall need to distinguish between the two when we come to consider calculus in the algebra of spacetime. Note that  $\nabla$  is a vector, so when acting on a vector  $A$  we have

$$\nabla A = \nabla \cdot A + \nabla \wedge A, \quad (2.68)$$

which shows that the vector derivative in 3D encompasses the divergence and curl, and that when it acts on a scalar it is the same as grad. By convention  $\nabla$  only operates on the object to its immediate right unless brackets are used to indicate otherwise, or over-stars or overdots are used to show what it operates on.

We can also define a more general multivector derivative. [1] Define the derivative with respect to  $X$  in the  $A$  ‘direction’ by

$$A * \partial_X F(X) \equiv \lim_{\tau \rightarrow 0} \frac{F(X + \tau P_X(A)) - F(X)}{\tau} \quad (2.69)$$

Here  $A * \partial_X$  is a single scalar operator and  $P_X(A)$  projects  $A$  into the grades in  $X$ . We now use a multivector basis  $e_K$  for the algebra under consideration. As when  $e_K$  was a vector we define the reciprocal frame  $e^K$  of multivectors so that  $\langle e^J e_K \rangle = \delta_K^J$ . We can now define the multivector derivative

$$\partial_X = e^K e_K * \partial_X. \quad (2.70)$$

A useful result that can be used to calculate more complicated results is

$$\partial_X \langle X A \rangle = e^K \langle e_K * \partial_X X A \rangle = e^K \langle P_X(e_K) A \rangle \quad (2.71)$$

$$= e^K \langle e_K P_X(A) \rangle = P_X(A). \quad (2.72)$$

Here are some examples of how the derivative works, where  $n$  is the space dimension:

$$\partial_a a = e^i e_i \cdot \partial_a a = e^i \cdot e_i = n \quad (2.73)$$

$$\partial_a a \cdot b = e^i (e_i \cdot \partial_a a) \cdot b = e^i e_i \cdot b = b \quad (2.74)$$

$$\partial_a b a = \partial_a (2a \cdot b - ab) = (2 - n)b. \quad (2.75)$$

$$\partial_X \langle X \tilde{X} \rangle = \tilde{X} + \dot{\partial}_X \langle X \tilde{X} \rangle = \tilde{X} + \dot{\partial}_X \langle \dot{X} \tilde{X} \rangle = 2\tilde{X} \quad (2.76)$$

## 2.8 Electromagnetism

Here we briefly show how GA can be useful in the description of electromagnetism. The Faraday tensor of conventional treatments,  $F_{\mu\nu}$ , is given by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (2.77)$$

where  $A_\nu$  is the vector potential. We can write this as

$$F_{\mu\nu} = (e_\nu \wedge e_\mu) \cdot (\nabla \wedge A) \quad (2.78)$$

as can be checked using the relation

$$(a \wedge b) \cdot (c \wedge d) = b \cdot ca \cdot d - b \cdot da \cdot c. \quad (2.79)$$

We can therefore equally well use the Faraday bivector  $F = \nabla \wedge A$  instead of an antisymmetric tensor. This is typical of translations into GA — a tensor is represented by some geometric object which naturally has the correct symmetry properties. This is equivalent to the use of the Faraday 2-form in differential geometry, however we can now use the unique features of GA to manipulate the bivector.

In a given  $\gamma_0$  frame we can split  $F$  into electric and magnetic parts,  $F = \mathbf{E} + I\mathbf{B}$ , where

$$\mathbf{E} = \frac{1}{2}(F - \gamma_0 F \gamma_0) \quad (2.80)$$

$$I\mathbf{B} = \frac{1}{2}(F + \gamma_0 F \gamma_0). \quad (2.81)$$

In the  $\gamma_0$  frame this gives a unique split of  $F$  into  $\mathbf{E}$  and  $\mathbf{B}$  since  $\mathbf{E}$  and  $I\mathbf{B}$  are orthogonal spacetime planes. We also need a 4-current  $J$  which is split

$$J\gamma_0 = \mathbf{J} + \rho. \quad (2.82)$$

To split  $\nabla$  we have to remember that it was defined in terms of a reciprocal frame, which gives

$$\gamma_0 \nabla = \partial_t + \nabla \quad (2.83)$$

It is now possible to write Maxwell's equations all in one go as

$$\nabla F = J. \quad (2.84)$$

We can check that this is correct by left multiplying by  $\gamma_0$ :

$$\gamma_0 \nabla F = (\partial_t + \nabla)(\mathbf{E} + I\mathbf{B}) \quad (2.85)$$

$$= \nabla \cdot \mathbf{E} + I\nabla \cdot \mathbf{B} + \frac{\partial \mathbf{E}}{\partial t} + I \frac{\partial \mathbf{B}}{\partial t} + \nabla \wedge \mathbf{E} + I\nabla \wedge \mathbf{B} \quad (2.86)$$

$$= \gamma_0 J = \rho - \mathbf{J} \quad (2.87)$$

The outer products can be converted to curls by using  $\nabla \times \mathbf{a} = -I\nabla \wedge \mathbf{a}$ . If we then equate parts of the same grade on both sides we recover the four Maxwell's equations, corresponding to the four grades of the 3-space algebra.

## 2.9 Linear functions and notation

Linear functions are very important in physics, usually in the form of tensors which describe various physical fields. We can write the action of a linear function explicitly, for example a vector valued function  $f$  acting on a vector can be written

$$a \rightarrow \underline{f}(a). \quad (2.88)$$

We use the underbar on any non-scalar linear function to emphasize that the result is not a scalar. Such linear functions are often represented with respect to a particular frame by matrices:

$$a_\mu \rightarrow f_\mu^\nu a_\nu \quad (2.89)$$

where  $f_\mu^\nu$  are the components of the function in a particular basis  $\{e_\mu\}$  given by

$$f_\mu^\nu = e_\mu \cdot \underline{f}(e^\nu). \quad (2.90)$$

In GA we prefer to maintain a frame-free notation. However the use of indices is still useful, so for a constant vector  $a$  we define

$$\underline{f}_a \equiv \underline{f}(a). \quad (2.91)$$

It is also sometimes useful to have a scalar representation of a function, in which case we define

$$f_{ab} \equiv a \cdot \underline{f}(b). \quad (2.92)$$

We drop the underbar to show that it is now a scalar (and not, for example, a vector valued function of two vectors). We can also write the component of a vector  $V$  in the direction  $a$  as  $V_a \equiv a \cdot V$ . This is consistent notation as  $V_a$  could equally well be viewed as a scalar valued function of the vector  $a$ . Clearly if we set  $a = e_i$ ,  $b = e_j$  we recover the components in that frame, however this is an ‘abstract index’ notation which does not need to reference a particular frame.

A common matrix operation is the transpose which in GA we call the *adjoint*. We write it with an overbar as  $\bar{f}_a$  and it is defined by

$$a \cdot \underline{f}(b) = \bar{f}(a) \cdot b \quad (2.93)$$

so that the scalar representation of  $\bar{f}(a)$  is  $b \cdot \bar{f}(a) = f_{ab}$ . If  $f_{ab} = f_{ba}$  the function is symmetric. It is when discussing this kind of symmetry property that the scalar index notation can be most transparent.

Contractions in GA can be formed by using the derivative:

$$\partial_a \cdot \underline{f}(a) = e_\nu \cdot \partial_a e^\nu \cdot \underline{f}(a) = e^\nu \cdot \underline{f}(e_\nu \cdot \partial_a a) = e^\nu \cdot \underline{f}(e_\nu). \quad (2.94)$$

This notation was used extensively in[5], however here we prefer to use a notation closer to tensor notation and write  $\partial_a$  as  $\gamma^a$ . In our index notation we define

$$\underline{f}^a \equiv \underline{f}(\gamma^a) \quad (2.95)$$

and the contraction can be written as

$$\gamma^a \cdot \underline{f}_a = f^a_a. \quad (2.96)$$

Similarly for a vector  $V$  we write  $V^a \equiv \gamma^a \cdot V$ . When we are using index notation it can be more consistent to write an arbitrary constant vector  $a$  in an equation as  $\gamma_a$ . If required

we can then replace the  $\gamma_a$  and  $\gamma^a$  with particular frame vectors  $e_i$  and  $e^i$  to obtain a representation in a particular frame. We therefore have the equivalent forms

$$\gamma^a \cdot \underline{f}_a \equiv \gamma^a \cdot \underline{f}(\gamma_a) \equiv \partial_a \cdot \underline{f}(a) = f^a{}_a = e^i \cdot \underline{f}(e_i). \quad (2.97)$$

and our previous definitions can be written

$$\gamma_a \cdot \underline{f}_b \equiv \gamma_b \cdot \bar{f}_a \quad f_{ab} \equiv \gamma_a \cdot \underline{f}_b \equiv \gamma_a \cdot \underline{f}(\gamma_b) \quad (2.98)$$

We may also wish to consider linear functions of other multivectors, not just vectors. Clearly it is not a problem to define such objects. However often we simply want to extend a function of a vector to act on other objects, and to this end we define the action on a blade by *outermorphism*:

$$\underline{f}(a_1 \wedge a_2 \wedge \dots \wedge a_n) \equiv \underline{f}(a_1) \wedge \underline{f}(a_2) \wedge \dots \wedge \underline{f}(a_n). \quad (2.99)$$

Then by linearity we can construct the function acting on any multivector. For example the determinant is just the volume scale factor of a linear function and is therefore given simply by

$$\det f = I^{-1} \underline{f}(I) \quad (2.100)$$

where  $I$  is a general pseudoscalar, the volume element for the space under consideration. Useful relations for the inverse functions are derived in [1] and given by

$$\underline{f}^{-1}(A) = \det f^{-1} \bar{f}(AI) I^{-1} \quad (2.101)$$

$$\bar{f}^{-1}(A) = \det f^{-1} I^{-1} \underline{f}(IA) \quad (2.102)$$

An important class of linear functions are tensors. These transform under rotations as

$$\underline{f}(a) \rightarrow R \underline{f}(\tilde{R}aR) \tilde{R}. \quad (2.103)$$

For example if  $\underline{f}(a)$  is determined by vectors  $V$  and  $X$  as  $\underline{f}(a) = a \cdot XV$  it transforms as

$$\underline{f}(a) \rightarrow (\tilde{R}aR) \cdot X RV \tilde{R} = a \cdot (RX \tilde{R}) RV \tilde{R} \quad (2.104)$$

which is what we would have obtained if we had written down  $\underline{f}(a)$  after the rotation.

To apply Lagrangian mechanics to general field theories we need to extend our geometric calculus to allow differentiation with respect to a linear function. [5] To do this we define the components of a multivector valued linear function of a vector  $X(a)$

$$X_{Ij} \equiv \langle e_I X(e_j) \rangle \quad (2.105)$$

and assemble the full derivative

$$\partial_{X(a)} \equiv a \cdot e_j e_I \partial_{X_{Ij}}. \quad (2.106)$$

The fundamental property is that

$$\partial_{X(a)}\langle X(b)M \rangle = a \cdot e_j e_I \partial_{X_{I_j}}(X_{I_j} b^j [P_X(M)]^I) \quad (2.107)$$

$$= a \cdot b e_I [P_X(M)]^I \quad (2.108)$$

$$= a \cdot b P_X(M). \quad (2.109)$$

We could easily extend the definition to linear functions of more than one argument. As an example we can calculate the derivative of the determinant of a vector valued function  $\underline{h}(a)$ . Using the general relation [1]

$$a \cdot (b \wedge c \wedge d \wedge \dots) = a \cdot b (c \wedge d \wedge \dots) - a \cdot c (b \wedge d \wedge \dots) + a \cdot d (b \wedge c \wedge \dots) - \dots \quad (2.110)$$

remembering that  $\underline{h}(A_r)$  is an outermorphism and using Leibnitz rule we have

$$\partial_{\underline{h}(a)}\langle \underline{h}(A_r)B_r \rangle = \underline{h}(a \cdot A_r) \cdot B_r. \quad (2.111)$$

Now we can calculate the derivative of the determinant:

$$\partial_{\underline{h}(a)} \det h = \partial_{\underline{h}(a)}\langle \underline{h}(I)I^{-1} \rangle \quad (2.112)$$

$$= \underline{h}(a \cdot I)I^{-1} \quad (2.113)$$

$$= \bar{h}^{-1}(a) \det h \quad (2.114)$$

where in the last line we have used the inverse (2.102). Similarly

$$\partial_{\bar{h}(a)} \det h = \underline{h}^{-1}(a) \det h. \quad (2.115)$$

## 2.10 Conclusions

As we have seen, GA provides a clear way for manipulating physical objects. The algebra of spacetime allows Lorentz transformations to be carried out by rotations, in a manner clearly analogous to rotations in 3-space. We have a co-ordinate free ‘abstract index’ notation that frees us from reference to a specific co-ordinate basis. GA is close in spirit to the language of differential forms; however forms lack the geometric product and therefore lack the power of GA.

In the subsequent chapters we consider some applications of GA and assume familiarity with the results in this chapter. When we consider electron scattering we shall find that the GA approach has much to offer. In the formulation of gauge theory gravity we also find some significant benefits. However as we move on to cosmological applications we mostly use the subset that is equivalent to traditional tensor manipulations. Ultimately, as when we consider perturbations in cosmology, we often reduce to a set of scalar equations that can be handled numerically by computer.

We believe that it makes sense to use the most appropriate tool to study a problem. When GA is optimal then use GA, when tensor notation is best we use that. Rather than dogmatically applying one set of notation to all problems we have tried to develop a notation that shows the connections between the different formulations and allows us to move between them as convenient.



## Chapter 3

# Electron Scattering in the Spacetime Algebra

The Spacetime Algebra provides an elegant language for studying the Dirac equation and helps illuminate the correspondence with classical theory. Cross-section calculations can be performed in an intuitive way following a method suggested by Hestenes [3]. The  $S$ -matrix is replaced with a scattering operator which rotates and dilates the initial states into the scattered states. We show how the method neatly handles spin dependence by allowing the scattering operator to become a function of the initial spin. When the operator is independent of spin we can provide manifestly spin-independent results. Spin basis states are not needed, and we do no spin sums, instead dealing with the spin orientation directly. We perform some example calculations for single electron scattering and briefly discuss more complicated cases in QED.

### 3.1 Introduction

The Dirac equation governs the quantum behaviour of particles with spin and can be represented mathematically in various ways. In Dirac's original formulation abstract 'gamma matrices' are introduced in a rather mysterious way. The gamma matrices satisfy a Clifford Algebra and have no direct physical interpretation. Here we use the Geometric Algebra formulation of the equation which is physically equivalent but rather less abstract and allows for a more physical interpretation. The complex spinors of the traditional approach are replaced by even multivectors that have a direct interpretation in terms of the probability density and spin orientation. In the GA formulation the Clifford Algebra structure is built into the algebra of spacetime and is no surprise.

The GA formulation of Dirac theory replaces the matrices of the conventional theory with multivectors, allowing a direct correspondence with the equation governing a classical spinning particle. We introduce the STA form of the Dirac equation, and show how the theory can be developed within the STA. Using the STA formulation Hestenes [3] has demonstrated an elegant method for performing cross-section calculations, which was further developed in [4]. We extend and clarify this work, handling spin-dependence in a

natural way.

Methods for calculating spinor cross-sections are well known, however these usually involve complicated abstract calculations with gamma matrices. In this chapter we show how to calculate cross-sections in a more transparent and intuitive way. Instead of using spin basis states, summing over spins and using spin projection operators, we instead incorporate the spin orientation directly. This greatly streamlines the calculation of spin dependent results, and makes it clear when results are independent of spin. We first consider single electron scattering, where our method is most naturally applied, and then briefly discuss multi-particle scattering.

## 3.2 The Dirac equation

The Hestenes STA form of the Dirac equation is entirely equivalent to the usual equation [20]. However the STA approach brings out the geometric structure, leading to more physically transparent calculations. Here we show that it is possible to arrive at the Dirac equation by quantizing a classical equation. This ‘derivation’ has the advantage that the observables are then clearly related to the classical parameters, and the geometric structure of the theory is brought out.

Our classical model will consist of a small spinning symmetric top with four velocity  $v$ . We can represent  $v$  as a boosted version of the lab frame time vector  $\gamma_0$ :

$$v = L\gamma_0\tilde{L} \quad (3.1)$$

where  $L$  is a boosting rotor. In this way the velocity can be represented by the rotor  $L$ . Similarly we can use a spatial rotor  $U$  to encode the spin plane as a rotation of some fixed reference plane. We write the rest spin of the top as

$$\hat{S}^0 = U\Sigma\tilde{U} \quad (3.2)$$

where  $\Sigma$  is some arbitrary constant reference bivector orthogonal to  $\gamma_0$  ( $\Sigma = \Sigma_3$  is often chosen). Since  $U$  is a spatial rotor it does not affect the  $\gamma_0$  direction so the momentum can be written

$$p = mR\gamma_0\tilde{R} \quad (3.3)$$

where  $R = LU$ . This equation for  $p$  squares to give  $p^2 = m^2$  which gives the Klein-Gordon equation on quantization. However the rotor equation contains much more information than the scalar equation given by its square.

The full rotor  $R = LU$  can be used to define the relativistic spin bivector by boosting up  $\hat{S}^0$

$$\hat{S} = L\hat{S}^0\tilde{L} = R\Sigma\tilde{R}. \quad (3.4)$$

As well as encoding the rotation of  $\Sigma$  into the spin plane the spatial rotor  $U$  can also include an arbitrary unobservable rotation in the reference plane  $\Sigma$ . The full rotor  $R$  therefore encodes everything about the four velocity and spin direction of the top as well some arbitrary unobservable rotation in the spin plane.



In the quantum version we wish to have probability densities. In the rest frame of the top this corresponds to some probability density  $\rho$  of finding it at each point. We want this to be the  $v \cdot J$  component of a four vector probability current  $J$ , with the lab frame probability density given by  $\gamma_0 \cdot J$ . We therefore define the four vector  $J = \rho v$  which can be written

$$J = \rho R \gamma_0 \tilde{R}. \quad (3.5)$$

We now wrap up  $\rho$  and  $R$  into a single even multivector  $\psi = \rho^{1/2} R$  so that

$$J = \psi \gamma_0 \tilde{\psi} \quad (3.6)$$

and the rest frame probability density is given by  $\rho = \psi \tilde{\psi}$ . We now want to put the equation for the momentum in terms of  $\psi$ . Multiplying the equation on the right by  $R$  we have

$$pR = mR\gamma_0 \quad (3.7)$$

$$\implies p\psi = m\psi\gamma_0. \quad (3.8)$$

This equation now contains all the ingredients for successful quantization. The usual procedure is to make the replacement  $p_\mu \rightarrow \hat{j}\nabla_\mu$ , so we get

$$\hat{j}\nabla\psi = m\psi\gamma_0 \quad (3.9)$$

as our form of the Dirac equation. For a plane wave  $\psi(x) = \psi e^{-\hat{j}p \cdot x}$  this just gives us back our classical equation, as expected.

There is a remaining ambiguity in what  $\hat{j}$  is. It could be a scalar imaginary, or could it be something more physical? Multiplication by  $\hat{j}$  should just affect the phase of the wave function, we don't want the spin or momentum to be affected. So for plane waves, writing  $\psi' = \hat{j}\psi$ , we want

$$S' = \psi' \Sigma \tilde{\psi}' = \psi \Sigma \tilde{\psi} \quad \text{and} \quad J' = \psi' \gamma_0 \tilde{\psi}' = \psi \gamma_0 \tilde{\psi}. \quad (3.10)$$

These can be satisfied if

$$\psi' \equiv \hat{j}\psi = \hat{S}\psi \quad \text{or} \quad \psi' \equiv \hat{j}\psi = \psi \Sigma, \quad (3.11)$$

and indeed for plane wave states these are equivalent since

$$\psi \Sigma = 1/\rho \psi \Sigma \tilde{\psi} \psi = \hat{S}\psi. \quad (3.12)$$

So the 'complex' phase factors of the form  $e^{\hat{j}\alpha}$  just encodes rotations in the spin plane — the rotations that were unobservable in the classical case.

### 3.3 Dirac theory

Having 'derived' the Dirac equation we now take that equation as given and see what it implies. For positive energy plane wave states all the classical results still hold. However we now have two sets of plane wave solutions,

$$\psi^{(+)} = u(p)e^{-\hat{j}p \cdot x} \quad \text{and} \quad \psi^{(-)} = v(p)e^{\hat{j}p \cdot x} \quad (3.13)$$

where

$$mu - pu\gamma_0 = 0 \quad \text{and} \quad mv + pv\gamma_0 = 0. \quad (3.14)$$

We therefore have positive and negative energy states, with energy projection operators given by

$$\Lambda_{\pm}(\psi) = \frac{1}{2m}(m\psi \pm p\psi\gamma_0). \quad (3.15)$$

Since  $\psi\tilde{\psi}$  reverses to itself it can only contain scalar and pseudoscalar parts and we define

$$\rho e^{I\beta} \equiv \psi\tilde{\psi} \quad (3.16)$$

where  $\beta$  and  $\rho$  are scalars. So the general form for  $\psi$  is now

$$\psi = \rho^{1/2} e^{I\beta/2} R. \quad (3.17)$$

In addition to encoding a rotation and a dilation, the spinor also contains a ‘ $\beta$ -factor’. This determines the ratio of particle to anti-particle solutions since

$$\Lambda_{\pm}(I\psi) = I\Lambda_{\mp}(\psi). \quad (3.18)$$

The transformation properties of  $\psi$  are inherited from its component rotor, so we have

$$\psi_L(x) = R\psi(\tilde{R}xR). \quad (3.19)$$

We call an element of the STA that transforms as a rotor a *spinor*.

The Dirac equation can be obtained from the Lagrangian

$$\mathcal{L} = \langle \hat{j}\nabla\psi\gamma_0\tilde{\psi} - m\psi\tilde{\psi} \rangle \quad (3.20)$$

by using the multivector form of the Euler-Lagrange equations [5, 21]. The Lagrangian is invariant under

$$\psi \rightarrow \psi e^{\hat{j}\theta} \quad (3.21)$$

corresponding to invariance under rotation in the spin plane. Using the multivector form of Noether’s theorem [21], we find the corresponding conserved probability current

$$J = \psi\gamma_0\tilde{\psi}, \quad (3.22)$$

in agreement with our classical definition. The Dirac equation ensures that  $\nabla \cdot J = 0$ .

### 3.4 Plane waves and basis states

Using the decomposition  $R = LU$  of a rotor into a spatial rotation  $U$  and a boost  $L$  we can write a spinor  $\psi$  as

$$\psi = \rho^{1/2} e^{I\beta/2} LU. \quad (3.23)$$

Consider a positive energy spinor  $u = \Lambda_+(u)$  and a negative energy spinor  $v = \Lambda_-(v)$ . If the particle is at rest we have

$$\gamma_0 u^0 \gamma_0 = u^0 \quad \text{and} \quad \gamma_0 v^0 \gamma_0 = -v^0 \quad (3.24)$$

which implies that

$$u^0 = \rho_u^{1/2} U_u \quad \text{and} \quad v^0 = \rho_v^{1/2} I U_v. \quad (3.25)$$

We can find the more general form by performing a boost to momentum  $p$ . The boost transforms  $m\gamma_0$  into the momentum  $p$ :

$$p = mL\gamma_0\tilde{L} \quad \Longrightarrow \quad pL\gamma_0 - mL = 0 \quad (3.26)$$

so that  $\Lambda_-(L) = 0$ . A solution is therefore of the form  $L = \Lambda_+(X)$ . Choosing  $X$  equal to a constant so that  $\tilde{L}L = 1$  we have

$$L = \frac{m + p\gamma_0}{\sqrt{2m(E + m)}} = \frac{E + m + \mathbf{p}}{\sqrt{2m(E + m)}}. \quad (3.27)$$

Normalizing so that  $\rho_u = \rho_v = 2m$  and performing the boost we get

$$u(p) = Lu^0 = \sqrt{E + m} \left( 1 + \frac{\mathbf{p}}{E + m} \right) U_u \quad (3.28)$$

$$v(p) = Lv^0 = I\sqrt{E + m} \left( 1 + \frac{\mathbf{p}}{E + m} \right) U_v. \quad (3.29)$$

In addition to the energy projection operators there are also the projection operators

$$\chi_{\pm}(\psi) = \frac{1}{2}(\psi \mp P\psi\Sigma) \quad (3.30)$$

where  $P$  is a bivector with  $P^2 = -1$ . For a state  $\psi$  satisfying  $\psi = \chi_{\pm}(\psi)$  we have

$$\psi = \mp P\psi\Sigma. \quad (3.31)$$

Multiplying on the right by  $\tilde{\psi}$  this gives  $\rho = \mp\rho P\hat{S}$  and so  $\hat{S} = \pm P$ . The projection operator therefore projects out parts corresponding the two spin orientations in the plane  $P$ . The spin projection operators commute with the energy projection operators

$$\Lambda_{\pm}(\psi) = \frac{1}{2m}(m\psi \pm p\psi\gamma_0) \quad (3.32)$$

since  $P \cdot p = 0$ . We can therefore split an arbitrary spinor (eight real components) into scalar and  $\hat{j}$  multiples of four basis states

$$u_1 = \chi_+(\Lambda_+(u_1)) \quad u_2 = \chi_-(\Lambda_+(u_2)) \quad (3.33)$$

$$v_1 = \chi_-(\Lambda_-(v_1)) \quad v_2 = \chi_+(\Lambda_-(v_2)) \quad (3.34)$$

With the normalization convention that  $u\tilde{u} = 2m$  the four basis states obey the orthogonality relations

$$\langle \tilde{u}_r u_s \rangle_S = 2m\delta_s^r \quad \langle \tilde{v}_r v_s \rangle_S = -2m\delta_s^r \quad (3.35)$$

$$\langle \tilde{u}_r v_s \rangle_S = 0 \quad \langle \tilde{v}_r u_s \rangle_S = 0 \quad (3.36)$$

where  $\langle A \rangle_S$  represents the  $\{1, \Sigma\}$  projection of  $A$ :

$$\langle A \rangle_S \equiv \langle A \rangle - \langle A \Sigma \rangle \Sigma. \quad (3.37)$$

By writing  $\psi$  as a sum over basis states it is easy to see that

$$\sum_r u_r \langle \tilde{u}_r \psi \rangle_S = p\psi\gamma_0 + m\psi \quad \text{and} \quad \sum_r v_r \langle \tilde{v}_r \psi \rangle_S = p\psi\gamma_0 - m\psi. \quad (3.38)$$

So we see that the usual basis state results of Dirac theory can be formulated in the STA approach. However we shall now develop the scattering theory largely without resort to basis states.

### 3.5 Feynman propagators

We now consider how to handle scattering from a vector potential  $A$ . This requires solutions of the minimally coupled Dirac equation which can be written

$$\hat{j}\nabla\psi\gamma_0 - m\psi = eA\psi\gamma_0 \quad (3.39)$$

where  $e = -|e|$  is the electron charge. We use a Green's function for this equation satisfying

$$\hat{j}\nabla_x S_F(x - x')\psi(x')\gamma_0 - mS_F(x - x')\psi(x') = \delta^4(x - x')\psi(x') \quad (3.40)$$

so that an integral solution can be found from

$$\psi(x) = \psi_i(x) + e \int d^4x' S_F(x - x')A(x')\psi(x')\gamma_0 \quad (3.41)$$

where  $\psi_i$  satisfies the free-particle equation. Taking the Fourier transform we have

$$pS_F(p)\psi\gamma_0 - mS_F(p)\psi = \psi \quad (3.42)$$

where

$$S_F(x - x') = \int \frac{d^4p}{(2\pi)^4} S_F(p) e^{-\hat{j}p \cdot (x - x')}. \quad (3.43)$$

Operating on both sides with the energy projection operator  $\Lambda_+$  we can solve for the momentum space Feynman propagator:

$$(p^2 - m^2)S_F(p)\psi = p\psi\gamma_0 + m\psi \quad (3.44)$$

$$\Rightarrow S_F(p)\psi = \frac{p\psi\gamma_0 + m\psi}{p^2 - m^2 + \hat{j}\epsilon}. \quad (3.45)$$

The  $\hat{j}\epsilon$  ensures that the contour integral is in the  $\Sigma$  plane and that it is causal—positive energy waves propagate into the future and negative energy waves into the past. Fourier transforming back and performing the integral over  $dE$  we get

$$S_F(x - x')\psi = -2m\hat{j} \int \frac{d^3\mathbf{p}}{2E_p(2\pi)^3} [\theta(t - t')\Lambda_+(\psi)e^{-\hat{j}p \cdot (x - x')} + \theta(t' - t)\Lambda_-(\psi)e^{\hat{j}p \cdot (x - x')}] \quad (3.46)$$

where  $E = +\sqrt{\mathbf{p}^2 + m^2}$ .

The photon propagator is the Greens' function for Maxwell's equations. In the Lorentz gauge  $\nabla \cdot A = 0$  we have  $\nabla^2 A = J$ , so the Greens' function must satisfy

$$\nabla_x^2 D_F(x - x') = \delta^4(x - x'). \quad (3.47)$$

Taking the Fourier transform we can solve for the Feynman propagator

$$D_F(p) = \frac{-1}{p^2 + \hat{j}\epsilon}. \quad (3.48)$$

### 3.6 Electron scattering

For scattering calculations we write the wavefunction as the sum of an incoming plane wave and a scattered beam,  $\psi = \psi_i + \psi_{\text{diff}}$ , where  $\psi_{\text{diff}}$  is the solution at asymptotically large times given by

$$\psi_{\text{diff}}(x) = -2m\hat{j}e \int d^4x' \int \frac{d^3\mathbf{p}}{2E_p(2\pi)^3} \Lambda_+ [A(x')\psi(x')\gamma_0] e^{-\hat{j}p \cdot (x-x')}. \quad (3.49)$$

This can be written as a sum over final states

$$\psi_{\text{diff}}(x) = \int \frac{d^3\mathbf{p}_f}{2E_f(2\pi)^3} \psi_f(x), \quad (3.50)$$

where the subscript on  $\psi_f(x)$  labels the final momentum and the final states are plane waves of the form

$$\psi_f(x) \equiv \psi_f e^{-\hat{j}p_f \cdot x} \equiv -\hat{j}e \int d^4x' [p_f A(x')\psi(x') + mA(x')\psi(x')\gamma_0] e^{-\hat{j}p_f \cdot (x-x')}. \quad (3.51)$$

With this definition the number of scattered particles is given by

$$\int d^3\mathbf{x} \gamma_0 \cdot J_{\text{diff}} = \int \frac{d^3\mathbf{p}_f}{2E_f(2\pi)^3} \left[ \frac{\gamma_0 \cdot J_f}{2E_f} \right] \equiv \int \frac{d^3\mathbf{p}_f}{2E_f(2\pi)^3} N_f \quad (3.52)$$

where we have defined the number density per Lorentz invariant phase space interval to be

$$N_f \equiv \frac{\gamma_0 \cdot J_f}{2E_f} = \frac{\gamma_0 \cdot (\psi_f \gamma_0 \tilde{\psi}_f)}{2E_f} = \frac{\rho_f}{2m}. \quad (3.53)$$

The Born series perturbative solution is generated by iterating (3.41). In the first order Born approximation this amounts to simply replacing  $\psi(x')$  by  $\psi_i(x')$ . For plane waves of particles we have

$$\psi(x) = \psi e^{-\hat{j}p \cdot x} \quad \text{and} \quad m\psi\gamma_0 = p\psi \quad (3.54)$$

so the final states become

$$\psi_f = -\hat{j}e \int d^4x' [p_f A(x') + A(x') p_i] \psi_i e^{\hat{j}q \cdot x'} \quad (3.55)$$

$$= -\hat{j}e [p_f A(q) + A(q) p_i] \psi_i \quad (3.56)$$

where  $q \equiv p_f - p_i$ .

More generally we define

$$\psi_f = S_{fi} \psi_i \quad (3.57)$$

where  $S_{fi}$  is the *scattering operator* which rotates and dilates the initial states into the final states. Here the  $f$  and  $i$  indices label the initial and final momenta and the initial spin, so in general  $S_{fi} = S_{fi}(p_f, p_i, \hat{S}_i)$ . However  $S_{fi}$  does not depend on the final spin—instead the final spin is determined from the initial spin by a rotation encoded in  $S_{fi}$ . Since  $S_{fi}$  consists of a rotation and dilation it is convenient to decompose it as

$$S_{fi} = \rho_{fi}^{1/2} R_{fi} \quad (3.58)$$

where  $R_{fi}$  is a rotor. There is no  $e^{I\beta}$  part since we have particles scattering to particles, not a mixture of particles and antiparticles. The cross-section will be determined by the  $\rho_{fi}$  factor, as detailed in the next section. The rotor  $R_{fi}$  rotates states with momentum  $p_i$  into states with momentum  $p_f$ . It also relates the initial and final spins by

$$\hat{S}_f = R_{fi} \hat{S}_i \tilde{R}_{fi} \quad (3.59)$$

so the rest spins are related by

$$\hat{S}_f^0 = \tilde{L}_f \hat{S}_f L_f = \tilde{L}_f R_{fi} \hat{S}_i \tilde{R}_{fi} L_f = \tilde{L}_f R_{fi} L_i \hat{S}_i^0 \tilde{L}_i \tilde{R}_{fi} L_f. \quad (3.60)$$

We therefore define the *rest spin scattering operator*

$$U_{fi} \equiv \tilde{L}_f R_{fi} L_i \quad (3.61)$$

so that

$$\hat{S}_f^0 = U_{fi} \hat{S}_i^0 \tilde{U}_{fi}. \quad (3.62)$$

The rest spin scattering operator and the cross-section contain all the information about scattering of states with momentum  $p_i$  and spin  $\hat{S}_i$  into states with momentum  $p_f$ .

The form of the external line Feynman propagator (3.46) ensures that  $S_{fi}$  is of the form

$$S_{fi} = -\hat{j}(p_f M + M p_i) \quad (3.63)$$

where in the Born approximation example  $M = eA(q)$ . However in general  $M$  can have some  $\hat{j}$ -dependence in which case we can write

$$S_{fi} \psi_i = -\hat{j}(p_f [M_r + \hat{j}M_j] + [M_r + \hat{j}M_j] p_i) \psi_i \quad (3.64)$$

where  $M_j$  and  $M_r$  are independent of  $\hat{j}$ . Using  $\hat{j}\psi_i = \psi_i\Sigma = \hat{S}_i\psi_i$  and the fact that  $\hat{S}_i$  and  $p_i$  commute this can be written

$$S_{fi} = -\hat{j}(p_f M + M p_i) \quad (3.65)$$

where

$$M = M_r + M_j \hat{S}_i \quad (3.66)$$

now depends on the initial spin. We can thus replace dependence on the ‘imaginary’  $\hat{j}$  with dependence on the spin bivector.

Using  $mL^2 = p\gamma_0$  we can obtain  $U_{fi}$  from

$$U_{fi} \propto L_f \gamma_0 M L_i + \tilde{L}_f M \gamma_0 \tilde{L}_i. \quad (3.67)$$

### 3.7 Positron scattering and pair annihilation

Adapting the above results to positron scattering is straightforward. We just consider a negative energy plane wave coming in from the future and scattering into the past, so  $\psi_i(x) = \psi_2 e^{\hat{j}p_2 x}$  and

$$S_{fi}\psi_i = -\hat{j}(-p_1 M \psi_i + M \psi_i \gamma_0) \quad (3.68)$$

where  $p_1$  is the incoming positron momentum and  $p_2$  is the outgoing momentum. This then gives

$$S_{fi} = \hat{j}(p_1 M + M p_2), \quad (3.69)$$

amounting to the substitution  $p_f \rightarrow -p_1, p_i \rightarrow -p_2$ .

The other case to consider is when the incoming electron gets scattered into the past, corresponding to pair annihilation. In this case we have

$$S_{fi} = -\hat{j}(-p_2 M + M p_1) \quad (3.70)$$

where  $p_1$  and  $p_2$  are the incoming momenta of the electron and positron respectively. In this case we can decompose  $S_{fi}$  as

$$S_{fi} = \rho_{fi}^{1/2} I R_{fi} \quad (3.71)$$

since  $S_{fi}$  must now contain a factor of  $I$  to map electrons into positrons. This also implies

$$S_{fi} \tilde{S}_{fi} = -\rho_{fi}. \quad (3.72)$$

### 3.8 Cross-sections

The scattering rate into the final states per unit volume per unit time is given by

$$W_{fi} = \frac{1}{VT} N_f = \frac{1}{VT} \frac{\gamma_0 \cdot J_f}{2E_f} = \frac{\rho_f}{2mVT} \quad (3.73)$$

where  $\rho_f$  is given simply by

$$\rho_f = |S_{fi}|^2 \rho_i = \rho_{fi} \rho_i. \quad (3.74)$$

Here we have defined

$$|S_{fi}|^2 \equiv |S_{fi} \tilde{S}_{fi}| = \pm S_{fi} \tilde{S}_{fi} \quad (3.75)$$

where plus sign corresponds to electron to electron and positron to positron scattering, the minus sign to electron-positron annihilation. The cross-section is defined as

$$d\sigma = \frac{W_{fi}}{\text{Target density} \times \text{Incident flux}}. \quad (3.76)$$

When  $S_{fi}$  is of the form

$$S_{fi} = -\hat{j}(2\pi)^4 \delta^4(P_f - P_i) T_{fi} \quad (3.77)$$

where the delta function ensures momentum conservation ( $P_f = P_i$ ) we have

$$|S_{fi}|^2 = VT(2\pi)^4 \delta^4(P_f - P_i) |T_{fi}|^2. \quad (3.78)$$

Working in the  $J_i$  frame the target density is just  $\rho_i$ , so writing the incident flux as  $\chi$  we have

$$d\sigma = \frac{1}{2m\chi} (2\pi)^4 \delta^4(P_f - P_i) |T_{fi}|^2. \quad (3.79)$$

Alternatively we may have elastic scattering with just energy conservation ( $E_f = E_i$ ) and

$$S_{fi} = -\hat{j}2\pi\delta(E_f - E_i) T_{fi}. \quad (3.80)$$

In this case

$$|S_{fi}|^2 = 2\pi T \delta(E_f - E_i) |T_{fi}|^2. \quad (3.81)$$

A target density of  $1/V$  and an incident flux of  $|\mathbf{J}_i| = \rho_i |\mathbf{p}_i|/m$  then gives

$$d\sigma = \frac{\pi}{|\mathbf{p}_i|} \delta(E_f - E_i) |T_{fi}|^2. \quad (3.82)$$

Above we have considered the total number of particles scattered. If we are interested in the final spin we can find it using the spin scattering operator. However we might also like to consider the cross-section when we only observe particles with final spins in a certain plane  $\hat{S}_o$  (where  $\hat{S}_o \cdot \mathbf{p}_f = 0$ ). This is particularly relevant in examples like electron-positron annihilation where  $\psi_f$  is actually an input state and we would like to calculate the cross-section for arbitrary initial spins.

The spin projection operators into the  $\hat{S}_o$  plane are

$$\chi_{\pm}(\psi) = \frac{1}{2}(\psi \mp \hat{S}_o \psi \Sigma) \quad (3.83)$$

and we are interested in scattering into

$$\chi_{\pm}(\psi_f) = \chi_{\pm}(S_{fi} \psi_i). \quad (3.84)$$



Now if  $S_{fi}$  is in the form (3.63) we have

$$\chi_{\pm}(S_{fi}\psi_i) = -\frac{1}{2} \left[ (p_f M + M p_i) \psi_i \boldsymbol{\Sigma} \pm \hat{S}_o (p_f M + M p_i) \psi_i \right] \quad (3.85)$$

$$= -\frac{1}{2} \left[ (p_f M + M p_i) \hat{S}_i \pm \hat{S}_o (p_f M + M p_i) \right] \psi_i \quad (3.86)$$

$$= -\frac{1}{2} \left[ p_f (M \hat{S}_i \pm \hat{S}_o M) + (M \hat{S}_i \pm \hat{S}_o M) p_i \right] \psi_i. \quad (3.87)$$

Defining  $\chi_{\pm}(\psi_f) = S_{fi}^{\pm} \psi_i$  the scattering rate will be proportional to  $\rho_{fi}^{\pm}$  given by

$$|S_{fi}^{\pm}|^2 = \left\langle (m^2 M + p_f M p_i) (\widetilde{M} \mp \hat{S}_i \widetilde{M} \hat{S}_o) \right\rangle. \quad (3.88)$$

If we sum over final spins the  $\hat{S}_o$  term cancels out and we get the expected result for the total  $\rho_{fi}$ :

$$|S_{fi}|^2 = \langle (p_f M + M p_i) (\widetilde{M} p_f + p_i \widetilde{M}) \rangle = 2 \langle m^2 M \widetilde{M} + p_f M p_i \widetilde{M} \rangle \quad (3.89)$$

### 3.9 Coulomb scattering

As our first simple example we consider the first Born approximation in electron Coulomb scattering where we have an external field given by

$$A(x) = \frac{-Ze}{4\pi|\mathbf{x}|} \gamma_0. \quad (3.90)$$

In the first Born approximation  $M$  is given by  $M = eA(q)$  where the Fourier transform of  $A(x)$  is

$$A(q) = -\frac{2\pi Ze}{\mathbf{q}^2} \delta(E_f - E_i) \gamma_0 \quad (3.91)$$

and  $q \cdot \gamma_0 = E_f - E_i$ . Writing

$$S_{fi} = -\hat{j} 2\pi \delta(E_f - E_i) T_{fi} \quad (3.92)$$

and using energy conservation we have

$$T_{fi} = -\frac{Ze^2}{\mathbf{q}^2} (2E + \mathbf{q}) \quad (3.93)$$

so that the formula for the cross-section becomes

$$d\sigma = \left( \frac{Ze^2}{\mathbf{q}^2} \right)^2 \frac{\pi}{|\mathbf{p}_i|} \delta(E_f - E_i) (4E^2 - \mathbf{q}^2) \frac{d^3 \mathbf{p}_f}{2E_f (2\pi)^3}. \quad (3.94)$$

Using  $d^3 \mathbf{p}_f = |\mathbf{p}_f| E_f dE_f d\Omega_f$  we recover the Mott cross-section

$$\left( \frac{d\sigma}{d\Omega_f} \right)_{\text{Mott}} = \frac{Z^2 \alpha^2}{\mathbf{q}^4} (4E^2 - \mathbf{q}^2) = \frac{Z^2 \alpha^2}{4\mathbf{p}^2 \beta^2 \sin^4(\theta/2)} (1 - \beta^2 \sin^2(\theta/2)), \quad (3.95)$$

where

$$\mathbf{q}^2 = (\mathbf{p}_f - \mathbf{p}_i)^2 = 2\mathbf{p}^2(1 - \cos \theta) \quad \text{and} \quad \beta = |\mathbf{p}|/E. \quad (3.96)$$

The derivation is manifestly independent of initial spin, so the cross-section is spin independent. Of course the final and initial spins will be related by the rest spin scattering operator  $U_{fi}$ , where

$$U_{fi} \propto L_f L_i + \tilde{L}_f \tilde{L}_i \propto (E + m)^2 + \mathbf{p}_f \mathbf{p}_i. \quad (3.97)$$

If  $U_{fi}$  rotates by an angle  $\delta$  in the  $\hat{B}$  plane ( $\hat{B}^2 = -1$ ) it is given by

$$U_{fi} = e^{\delta \hat{B}/2} = \cos(\delta/2) + \hat{B} \sin(\delta/2). \quad (3.98)$$

So we see that the rotation is in the  $\mathbf{p}_f \wedge \mathbf{p}_i$  plane and by an angle  $\delta$  given by

$$\tan(\delta/2) = \frac{|\langle U_{fi} \rangle_2|}{\langle U_{fi} \rangle} = \frac{|\mathbf{p}_f \wedge \mathbf{p}_i|}{(E + m)^2 + \mathbf{p}_f \cdot \mathbf{p}_i} = \frac{\sin \theta}{(E + m)/(E - m) + \cos \theta}. \quad (3.99)$$

Similar derivations of these result using the STA approach have been given before [3, 22].

### 3.10 Compton scattering

In Compton scattering one electron interacts with two photons. We can therefore apply the above formalism in a somewhat heuristic way by using plane waves to represent the potentials of the two photons. There are two Feynman diagrams to consider which give two terms of the form

$$M_{12} = e^2 \int d^4 x' \int d^4 x'' \int \frac{d^4 p}{(2\pi)^4} A_1(x') \frac{p A_2(x'') + A_2(x'') p_i}{p^2 - m^2 + \hat{j}\epsilon} e^{\hat{j}x' \cdot (p_f - p)} e^{\hat{j}x'' \cdot (p - p_i)} \quad (3.100)$$

where

$$A(x) = \epsilon e^{\mp j k \cdot x} \quad (3.101)$$

is different at each vertex and  $\epsilon^2 = -1$ . Performing the integrations and summing the two contributions we have

$$M = e^2 (2\pi)^4 \delta^4(p_f + k_f - p_i - k_i) \left[ \epsilon_f \frac{(p_i + k_i) \epsilon_i + \epsilon_i p_i}{2k_i \cdot p_i} + \epsilon_i \frac{(p_i - k_f) \epsilon_f + \epsilon_f p_i}{-2p_i \cdot k_f} \right]. \quad (3.102)$$

Choosing  $p_i \cdot \epsilon_i = p_i \cdot \epsilon_f = 0$  this is simply

$$M = e^2 (2\pi)^4 \delta^4(p_f + k_f - p_i - k_i) \left( \frac{\epsilon_f h_i \epsilon_i}{2k_i \cdot p_i} + \frac{\epsilon_i k_f \epsilon_f}{2p_i \cdot k_f} \right). \quad (3.103)$$

Writing

$$S_{fi} = -\hat{j} (2\pi)^4 \delta^4(p_f + k_f - p_i - k_i) T_{fi} \quad (3.104)$$

and using (3.89) we then have

$$|T_{fi}|^2 = e^4 \left\langle \frac{m^2 \epsilon_f k_i \epsilon_i \epsilon_f k_f \epsilon_i + p_f \epsilon_f k_i \epsilon_i p_i \epsilon_f k_f \epsilon_f}{k_i \cdot p_i k_f \cdot p_i} + \frac{p_f \epsilon_f k_i \epsilon_i p_i \epsilon_i k_i \epsilon_f}{2(k_i \cdot p_i)^2} + \frac{p_f \epsilon_i k_f \epsilon_f p_i \epsilon_f k_f \epsilon_i}{2(k_f \cdot p_i)^2} \right\rangle. \quad (3.105)$$

The identities we need to calculate are now the same as in the traditional approach, only now we know that the result is independent of initial spin since we have not done a spin sum. Using momentum conservation we know

$$p_f + k_f = p_i + k_i \quad k_f \cdot p_i = k_i \cdot p_f \quad p_i \cdot k_i = p_f \cdot k_f. \quad (3.106)$$

Applying these the result becomes, after some work,

$$|T_{fi}|^2 = e^4 \left[ 4(\epsilon_i \cdot \epsilon_f)^2 - 2 + \frac{p_i \cdot k_f}{p_i \cdot k_i} + \frac{p_i \cdot k_i}{p_i \cdot k_f} \right]. \quad (3.107)$$

To calculate the cross-section we work in the frame where the electron is initially at rest ( $p_i = m\gamma_0$ ). The incoming photon flux is  $2k_i^0$  so we have

$$d\sigma = (2\pi)^4 \delta^4(p_f + k_f - p_i - k_i) \frac{|T_{fi}|^2}{2m2k_i^0} \frac{d^3k_f}{2k_f^0(2\pi)^3} \frac{d^3p_f}{2E_f(2\pi)^3}. \quad (3.108)$$

Now

$$\int d^3p_f d^3k_f \delta^4(p_f + k_f - p_i - k_i) = (k_f^0)^2 \frac{E_f k_f^0}{m k_i^0} d\Omega, \quad (3.109)$$

where we have done the integral over the final electron's momentum since we are primarily interested in the scattering of the photon. In the lab frame the result is therefore

$$\frac{d\sigma}{d\Omega} = \left( \frac{k_f}{k_i} \right)^2 \frac{|T_{fi}|^2}{4m^2(4\pi)^2} \quad (3.110)$$

$$= \frac{\alpha^2}{4m^2} \left( \frac{k_f}{k_i} \right)^2 \left[ \frac{k_f}{k_i} + \frac{k_i}{k_f} + 4(\epsilon_f \cdot \epsilon_i)^2 - 2 \right] \quad (3.111)$$

in agreement with the Klein-Nishina formula. Again, the difference is that this derivation applies regardless of the initial electron spin. Of course if we had used circularly polarized photons we would have introduced some  $\hat{j}$ -dependence and the result would have become spin-dependent.

### 3.11 Pair annihilation

A process closely related to Compton scattering is electron-positron annihilation. We just have to take account of the fact that the 'out' state is a positron so the final states are of the form

$$\psi_f(x) \equiv \psi_f e^{\hat{j}p_f \cdot x} \equiv -\hat{j}e \int d^4x' [-p_f A(x') \psi(x') + mA(x') \psi(x') \gamma_0] e^{\hat{j}p_f \cdot (x-x')}. \quad (3.112)$$

Writing

$$S_{fi} = -\hat{j}(-p_f M + M p_i) \quad (3.113)$$

we have two terms of the form

$$M_{12} = e^2 \int d^4 x' \int d^4 x'' \int \frac{d^4 p}{(2\pi)^4} A_1(x') \frac{p A_2(x'') + A_2(x'') p_i}{p^2 - m^2 + \hat{j}\epsilon} e^{-\hat{j}x' \cdot (p_f + p)} e^{\hat{j}x'' \cdot (p - p_i)} \quad (3.114)$$

where

$$A(x) = \epsilon e^{\hat{j}k \cdot x} \quad (3.115)$$

is different at each vertex and  $\epsilon^2 = -1$ . As for Compton scattering we now choose  $p_i \cdot \epsilon_1 = p_i \cdot \epsilon_2 = 0$  and sum the two contributions to get

$$M = e^2 (2\pi)^4 \delta^4(p_f + p_i - k_1 - k_2) \left( \frac{\epsilon_2 k_1 \epsilon_1}{2k_1 \cdot p_i} + \frac{\epsilon_1 k_2 \epsilon_2}{2p_i \cdot k_2} \right). \quad (3.116)$$

For general positron and electron spins we should use (3.88) to calculate the cross-section. However if either  $\psi_f$  or  $\psi_i$  are unpolarized the spin dependence will cancel out and the average just introduces a factor of two into equation (3.89). In this case  $|T_{fi}|^2$  is obtained from the Compton case by the substitution  $p_f \rightarrow -p_f$  and  $k_i \rightarrow -k_1$ , and an overall sign change because  $T_{fi} \tilde{T}_{fi} < 0$  in this case:

$$|T_{fi}|^2 = -\frac{e^4}{2} \left[ 4(\epsilon_i \cdot \epsilon_f)^2 - 2 - \frac{p_i \cdot k_2}{p_i \cdot k_1} - \frac{p_i \cdot k_1}{p_i \cdot k_2} \right]. \quad (3.117)$$

To get the cross-section just divide by flux factors and perform the integral as usual.

### 3.12 Second order Coulomb scattering

Second order Coulomb scattering is interesting as it is spin-dependent and so provides a good testing ground for our calculation techniques. To avoid problems with divergent integrals the potential is replaced with the screened potential

$$A(x) = -\frac{e^{-\lambda|\mathbf{x}|} Z e}{4\pi|\mathbf{x}|} \gamma_0 \quad (3.118)$$

and the Coulomb result found in the limit  $\lambda$  goes to zero [23, 24]. For this potential the first order analysis above can be applied with  $M$  given by

$$eA(q) = -\frac{2\pi Z e^2}{\lambda^2 + \mathbf{q}^2} \delta(E_f - E_i) \gamma_0. \quad (3.119)$$

To iterate to second order (3.51) is used, with the substitution

$$\psi(x') = \psi_i e^{-\hat{j}p_i \cdot x'} + e \int d^4 x'' \int \frac{d^4 k}{(2\pi)^4} \frac{k A(x'') + A(x'') p_i}{k^2 - m^2 + \hat{j}\epsilon} \psi_i e^{\hat{j}x'' \cdot (k - p_i)} e^{-\hat{j}k \cdot x'} \quad (3.120)$$

giving the extra contribution to  $M$

$$M' = e^2 \int d^4 x' \int d^4 x'' \int \frac{d^4 k}{(2\pi)^4} A(x') \frac{kA(x'') + A(x'')p_i}{k^2 - m^2 + \hat{j}\epsilon} e^{\hat{j}x' \cdot (p_f - k)} e^{\hat{j}x'' \cdot (k - p_i)}. \quad (3.121)$$

Carrying out the  $x'$  and  $x''$  integrations and using one of the resultant  $\delta$ -functions we have

$$M = 2\pi\delta(E_f - E_i)M_T \quad (3.122)$$

where the extra contribution to  $M_T$  is

$$M'_T = e^2 \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{A_0(p_f - k)A_0(k - p_i)}{k^2 - m^2 + \hat{j}\epsilon} \gamma_0 [k\gamma_0 + \gamma_0 p_i] \quad (3.123)$$

and

$$A_0(p) = \int d^3 \mathbf{x} e^{-\mathbf{p} \cdot \mathbf{x}} \gamma_0 \cdot A(\mathbf{x}) = \frac{-Ze}{\lambda^2 + \mathbf{p}^2}. \quad (3.124)$$

Using

$$k^2 - m^2 = \mathbf{p}_i^2 - \mathbf{k}^2 \quad (3.125)$$

and the integrals

$$I_1 + \frac{1}{2}(\mathbf{p}_i + \mathbf{p}_f)I_2 = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{1 + \mathbf{k}}{[(\mathbf{p}_f - \mathbf{k})^2 + \lambda^2][(\mathbf{p}_i - \mathbf{k})^2 + \lambda^2](\mathbf{p}_i^2 - \mathbf{k}^2 + \hat{j}\epsilon)} \quad (3.126)$$

we have

$$M'_T = Z^2 e^4 [\gamma_0 \frac{1}{2}(\mathbf{p}_i + \mathbf{p}_f)I_2 + (p_i + \gamma_0 E)I_1]. \quad (3.127)$$

In the limit  $\lambda \rightarrow 0$  our total  $M_T$  to second order is therefore

$$M_T = \frac{-Ze^2}{\mathbf{q}^2} \gamma_0 + Z^2 e^4 [(E\gamma_0 - \frac{1}{2}[p_f + p_i])I_2 + (p_i + \gamma_0 E)I_1] \quad (3.128)$$

where the integrals are [24]

$$I_1 = \frac{-\hat{j}}{16\pi|\mathbf{p}|^3 \sin^2(\theta/2)} \ln \frac{2|\mathbf{p}| \sin(\theta/2)}{\lambda} \quad (3.129)$$

$$I_2 = \frac{1}{16\pi|\mathbf{p}|^3 \cos^2(\theta/2)} \left\{ \frac{\pi[\sin(\theta/2) - 1]}{2 \sin^2(\theta/2)} - \hat{j} \ln \frac{\lambda}{2|\mathbf{p}|} \right\} + \frac{I_1}{\cos^2(\theta/2)}. \quad (3.130)$$

We see that  $M$  has some  $\hat{j}$  dependence, so writing  $I_1 = (A + C)\hat{j}$  and  $I_2 = B + C\hat{j}$  where  $A$ ,  $B$ , and  $C$  are scalars, and replacing the  $\hat{j}$ -dependence with  $\hat{S}_i$ -dependence, this becomes

$$M_T = \gamma_0 \left[ -\frac{Ze^2}{\mathbf{q}^2} + EZ^2 e^4 \left\{ B + (2C + A)\hat{S}_i \right\} \right] + Z^2 e^4 \left[ p_i(A\hat{S}_i - B) - \frac{1}{2}q(B + C\hat{S}_i) \right]. \quad (3.131)$$

The term proportional to  $q$  does not contribute to  $T_{fi}$ . Using

$$p_f p_i + m^2 = E(2E + \mathbf{q}) - \mathbf{p}^2 - \mathbf{p}_f \mathbf{p}_i \quad (3.132)$$

we have

$$T_{fi} = (2E + \mathbf{q}) \left[ -\frac{Ze^2}{\mathbf{q}^2} + 2EZ^2e^4(A + C)\hat{S}_i \right] + Z^2e^4(\mathbf{p}^2 + \mathbf{p}_f\mathbf{p}_i)(B - A\hat{S}_i). \quad (3.133)$$

Keeping terms up to  $\alpha^3$  the cross-section is governed by

$$|T_{fi}|^2 = (4E^2 - \mathbf{q}^2) \frac{Z^2e^4}{\mathbf{q}^4} - \frac{4Z^3e^6}{\mathbf{q}^2} \left[ EB(\mathbf{p}^2 + \mathbf{p}_f\mathbf{p}_i) + mA(\mathbf{p}_i \wedge \mathbf{p}_f) \cdot \hat{S}_i^0 \right] \quad (3.134)$$

where  $\hat{S}_i^0$  is the initial rest spin. As expected the divergent parts of the integrals have cancelled out, and we are only left with the finite terms  $B$  and

$$A = \frac{\ln \sin(\theta/2)}{16\pi|\mathbf{p}|^3 \cos^2(\theta/2)}. \quad (3.135)$$

The cross-section for unpolarized scattering is found by averaging over the initial spin. This gives the spin-independent part of the cross-section since the spin dependent part averages to zero. The  $\alpha^3$  contribution is therefore

$$\frac{d\sigma'}{d\Omega_f} = -\frac{4Z^3e^6BE}{4(2\pi)^2\mathbf{q}^2}(\mathbf{p}^2 + \mathbf{p}_f\mathbf{p}_i) \quad (3.136)$$

$$= \frac{\pi\alpha^3Z^3E[1 - \sin(\theta/2)]}{4|\mathbf{p}|^3 \sin^3(\theta/2)}. \quad (3.137)$$

Hence the unpolarized cross-section, including the second Born approximation but ignoring radiative corrections, is

$$\frac{d\sigma}{d\Omega_f} = \left( \frac{d\sigma}{d\Omega_f} \right)_{\text{Mott}} \left\{ 1 + Z\alpha\pi \frac{\beta \sin(\theta/2)[1 - \sin(\theta/2)]}{1 - \beta \sin^2(\theta/2)} \right\} \quad (3.138)$$

in agreement with the result obtained by Dalitz [23] using the conventional matrices and spin-sums approach.

### 3.13 Spin dependence and double scattering

As an example of handling spin dependence we can work out the asymmetry parameter for double scattering from a Coulomb potential. The idea is that since the second order correction to Coulomb scattering is spin dependent the scattered beam will be partially polarized even with an unpolarized incident beam. The scattered beam can then impinge on a second target, which leads to an observable asymmetry in the scattered intensity. The setup is shown in Figure 3.1. The asymmetry was first worked out by Mott [25, 26].

The first thing we need to know is the spin after the first scattering. This is given by

$$\hat{S}_f = R_{fi}\hat{S}_i\tilde{R}_{fi} \quad (3.139)$$

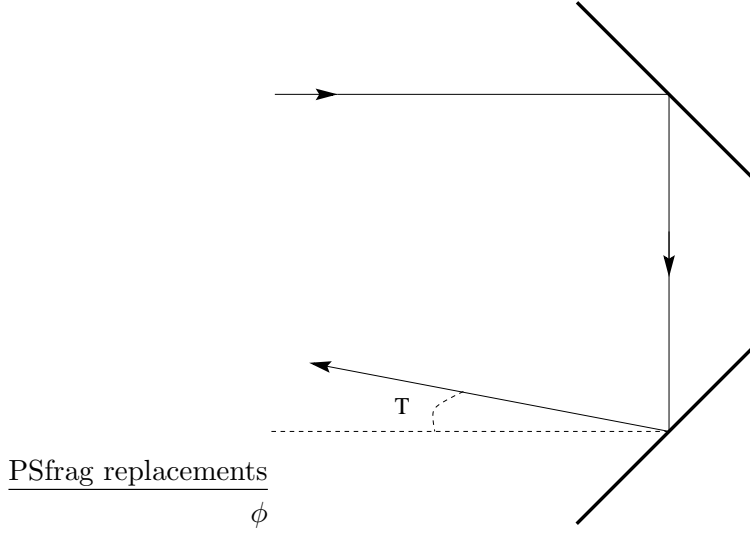


Figure 3.1: A double scattering experiment. An incoming unpolarized beam scatters off one target and then off a second target. The dependence on  $\phi$ , the angle of the second scattering to the plane of the first scattering, is determined by the asymmetry parameter  $\delta$ .

so we have

$$\hat{S}_f \propto T_{fi} \hat{S}_i \tilde{T}_{fi} = \frac{Z^2 e^4}{q^4} (2E + \mathbf{q}) \hat{S}_i (2E - \mathbf{q}) - \frac{2Z^3 e^6 A}{q^2} \langle (\mathbf{p}^2 + \mathbf{p}_f \mathbf{p}_i) (2E - \mathbf{q}) \rangle_2 \quad (3.140)$$

where we have only kept the lowest order terms in the spin dependent and spin-independent parts. We now define  $S^0$  to be the polarization in the plane  $\hat{S}^0$ . This is just a bivector in the plane of  $\hat{S}^0$  with modulus equal to the polarization of the beam. Since the incoming beam is taken to be unpolarized the resultant polarization plane will be given by the spin-independent part of  $\hat{S}_f$  deboosted to rest. To get the polarization we then just divide by the magnitude of the spin-dependent part:

$$S_f^0 = -\frac{2Ze^2 q^2 A}{(4E^2 - q^2)} \tilde{L}_f \langle (\mathbf{p}^2 + \mathbf{p}_f \mathbf{p}_i) (2E - \mathbf{q}) \rangle_2 L_f \quad (3.141)$$

$$= \frac{2Ze^2 q^2 A}{(4E^2 - q^2)} 2m \mathbf{p}_i \wedge \mathbf{p}_f. \quad (3.142)$$

The spin-dependent part of the cross-section for the second scattering is then given by

$$\left( \frac{d\sigma}{d\Omega_f} \right)_{\text{spin}} = -\frac{4Z^3 e^6 m A_2}{q_2^2 (2\pi)^2} (\mathbf{p}_f \wedge \mathbf{p}_2) \cdot S_f^0 \quad (3.143)$$

$$= -\frac{64(2\pi)^2 Z^4 \alpha^4 q_1^2 m^2 A_1 A_2}{q_2^2 (4E^2 - q_1^2)} (\mathbf{p}_f \wedge \mathbf{p}_2) \cdot (\mathbf{p}_i \wedge \mathbf{p}_f) \quad (3.144)$$

where the 1 and 2 subscripts refer to the first and second scattering respectively (e.g.  $\mathbf{q}_2 = \mathbf{p}_2 - \mathbf{p}_f$ ). We see that the asymmetry will depend on the cosine of the angle  $\phi$  between the

$\mathbf{p}_f \wedge \mathbf{p}_2$  and  $\mathbf{p}_i \wedge \mathbf{p}_f$  planes. The asymmetry parameter  $\delta$  is defined so that the final intensity depends on  $\phi$  through the factor

$$1 + \delta \cos \phi. \quad (3.145)$$

In the case where  $\mathbf{p}_i \cdot \mathbf{p}_f = \mathbf{p}_f \cdot \mathbf{p}_2 = 0$  ( $\mathbf{p}_i \cdot \mathbf{p}_2 = -\mathbf{p}^2 \cos \phi$ ) we find that the first non-zero contribution to the asymmetry factor is

$$\delta = \frac{64(2\pi)^2 Z^4 \alpha^4 m^2 A^2}{(4E^2 - \mathbf{q}^2)} \mathbf{p}^4 \frac{\mathbf{q}^4}{Z^2 \alpha^2 (4E^2 - \mathbf{q}^2)} \quad (3.146)$$

$$= Z^2 \alpha^2 (\ln 2)^2 \frac{\beta^2 (1 - \beta^2)}{(2 - \beta^2)^2} \quad (3.147)$$

in agreement with the answer quoted by Dalitz [23]. It is of course only the first approximation, and for large  $Z$  nuclei higher order corrections will be far from negligible.

### 3.14 The partial spin-sum approach

The above formalism seems to work well for single particle scattering. Here we show how we can adapt a more traditional approach in more complicated cases, demonstrating the flexibility of the STA formalism. The scattering operator approach could equally well be used in the more complicated case, as we show below.

We use the two basis states  $u_r$  to write (3.51) as

$$\psi_f = -\hat{j}e \int d^4x' \sum_r u_r(p_f) \langle \tilde{u}_r(p_f) A(x') \psi_i \gamma_0 \rangle_S e^{\hat{j}p_f \cdot x'} \quad (3.148)$$

$$= \sum_r u_r(p_f) S_{fi}^r \quad (3.149)$$

where  $S_{fi}^r$  is the traditional  $S$ -matrix. The total number density per Lorentz invariant phase space interval is then

$$N_f = \sum_r |S_{fi}^r|^2. \quad (3.150)$$

As an example we consider electron-muon scattering (following, for example, Bjorken and Drell [27]) in which  $A$  is given by

$$A(x) = \int d^4x' D_F(x - x') J(x') \quad (3.151)$$

and  $J(x')$  is the ‘complex’ conserved current given by

$$J_a = e \langle \tilde{u}_s \gamma_a \psi_2 \gamma_0 \rangle_S. \quad (3.152)$$

Defining  $T^{rs}$  as usual we have

$$T^{rs} = -\frac{e^2}{q^2} \langle \tilde{u}_r \gamma_a \psi_1 \gamma_0 \rangle_S \langle \tilde{u}_s \gamma^a \psi_2 \gamma_0 \rangle_S. \quad (3.153)$$



where  $q = p'_1 - p_1 = p_2 - p'_2$  and dashed variables correspond to the final states. Summing over  $r$  and  $s$

$$|T|^2 = \frac{e^4}{q^4} \langle \gamma_0 \tilde{\psi}_1 \gamma^b (p'_1 \gamma^a + \gamma^a p_1) \psi_1 \rangle_S \langle \gamma_0 \tilde{\psi}_2 \gamma_b (p'_2 \gamma_a + \gamma_a p_2) \psi_2 \rangle_S \quad (3.154)$$

$$= \frac{2e^4 \rho_1 \rho_2}{m_1 m_2 q^4} \left[ \langle \gamma^b (p'_1 \gamma^a + \gamma^a p_1) p_1 \rangle \langle \gamma_b (p'_2 \gamma_a + \gamma_a p_2) p_2 \rangle \right. \quad (3.155)$$

$$\left. - \langle \gamma^b (p'_1 \gamma^a + \gamma^a p_1) p_1 \hat{S}_1 \rangle \langle \gamma_b (p'_2 \gamma_a + \gamma_a p_2) p_2 \hat{S}_2 \rangle \right] \quad (3.156)$$

$$= \frac{2e^4 \rho_1 \rho_2}{m_1 m_2 q^4} \left[ p'_1 \cdot p'_2 p_1 \cdot p_2 + p'_2 \cdot p_1 p_2 \cdot p'_1 - m_1^2 p'_2 \cdot p_2 + m_2^2 p'_1 \cdot p_1 + 2m_2^2 m_1^2 \right. \quad (3.157)$$

$$\left. - [q \cdot (\hat{S}_1 \wedge p_1)] \cdot [q \cdot (\hat{S}_2 \wedge p_2)] \right]. \quad (3.158)$$

This approach differs from the normal one in that we have only done one spin sum over the final spins. We can therefore explicitly retain information about the initial spins, and calculations that are spin-independent will be manifestly so. Spin averaging simply amounts to removing spin dependent terms in the cross-section.

The same result could be obtained using the scattering operator approach using

$$M = e D_F \gamma^a J_a \quad (3.159)$$

and summing over the final spin of the other particle. One ends up with exactly the same equation. However the scattering operator approach may be better for calculating spin effects. If we are interested in the spin dependence of a particular fermion line the scattering operator approach works well once we have summed over the spins of the other particles. For example we can calculate the final spin and polarization in the same way as we did for Coulomb scattering. In this approach we still have to perform a spin sum, but only over the spins of the other particles. We could of course introduce spin projection operators to single out particular spins of the other particles if necessary.

### 3.15 Conclusions

We have seen how Hestenes' STA formulation of Dirac theory provides a useful and elegant method of performing cross-section calculations. Spin is handled in a simple manner, and the logic of calculating cross-sections is simplified considerably. We do not perform unnecessary spin sums and spin dependence is manifest in the spin bivector dependence of the scattering operator. It's a simple matter to calculate spin precessions, polarizations and spin dependent results, and the results are automatically expressed in terms of physical spin bivectors and the other scattering parameters. We can perform unpolarized calculations simply by averaging over spins.

In the multiparticle case things are more complicated. We do not have a neat method for performing arbitrary spin dependent calculations, and still have to resort to spin sums over terms involving complex conserved currents. However we can still write down a scattering

operator for any given fermion line, retaining the benefits of the scattering operator for calculations involving the spin of the particle.

## Chapter 4

# Gauge Theories of Gravity

In this chapter we review the formulation of gauge theories of gravity and consider some extensions. We study a formulation using Geometric Algebra, relate it General Relativity and other formulations, and review the key features. We then extend the gauged symmetries to include local scale invariance and discuss the properties of a scale invariant action. The field equations are derived and gauge fixing discussed. By gauge fixing appropriately we recover standard (non-scale invariant) gauge theory gravity with the addition of a massive vector field that only couples to scalar fields. We compare our approach to corresponding theories in GR and mention possible further extensions.

### 4.1 Introduction

It is now many years since Einstein formulated his General Theory of Relativity (GR), and we now have convincing experimental confirmation of some its key predictions. The theory asserts that spacetime is curved and that the curvature is determined by the stress-energy tensor and boundary conditions. GR is a beautiful geometrical theory with much theoretical attractiveness. However it does have its drawbacks. For example GR only works with a symmetric stress-energy tensor, so it is not immediately clear what happens when matter with spin is present (which has an antisymmetric stress-energy tensor). The problems with quantizing the theory are also well documented [28].

In order to make GR consistent with matter with spin one can incorporate a torsion field. There are many ways of doing this so that the theory is consistent with GR when the torsion can be neglected [29–32]. The theories we study here are those based on gauge symmetries, which we collectively call Gauge Theory Gravity (GTG). Such theories have been formulated numerous times in slightly different ways and have been called ECKS, Poincaré, Einstein-Cartan, or  $U_4$  spin-torsion theories [31–35]. Here we base our work on GTG as formulated in Geometric Algebra by Lasenby *et al.* [5]. The essential symmetries to gauge are invariance under local rotations (Lorentz transformations) and displacements (diffeomorphisms). We refer to a theory with these symmetries as a Poincaré gauge theory. With an appropriate choice of action the Poincaré gauge theory reproduces the results of GR for all the standard tests, but also incorporates torsion in a natural manner. In terms

of GR this essentially amounts to replacing the connection defined in terms of the metric with a free field that has a field equation determined by varying an action.

After reviewing GTG, relating our formulation and notation to other works and reviewing some key results, we move on to consider extending the gauge symmetries. The most natural extension is to consider a local scaling symmetry. This then requires the construction of a fully locally scale invariant theory.

The idea of a locally scale invariant theory of gravity goes back a long way. Shortly after Einstein's theory was developed Weyl proposed an extension of Riemannian geometry in which vectors not only change direction on parallel displacement around a closed loop but also change their length. Weyl's non-Riemannian geometry was the first example of a gauge theory, the gauge group being local scale transformations. To enforce this new gauge symmetry one has to introduce a vector gauge field which Weyl identified with the vector potential of electromagnetism. Weyl's theory was a bold, though ultimately unsuccessful attempt to provide a geometric picture of electromagnetism in the same way that general relativity (GR) provided a geometric picture of gravity. The theory was rejected because it seemed to conflict with the absolute length scales appearing in quantum phenomena.

Dirac [36] took up Weyl's theme as a way of implementing his Large Numbers Hypothesis. He thought the gravitational constant could change with time in order to keep the dimensionless number formed from fundamental constants approximately equal to the age of the universe expressed in atomic units. To this end his gravitational Lagrangian coupled a scalar field to the Ricci scalar, forming a correctly scale invariant Lagrangian.

We now know that local scale invariance is not the symmetry that generates electromagnetism; however that does not mean that a scale gauge symmetry should be rejected out of hand. The subject has since been studied by various authors, for instance see [37–46].

In the gauge theory approach to gravity all observables must be gauge invariant, in particular they must be scale gauge invariant. The result of a measurement is a dimensionless number, we do not measure dimensionful quantities. Since observables are merely ratios of dimensionful quantities there is a global invariance under rescaling of all lengths (and masses, etc, appropriately). One could equally well regard this as invariance under change of units. Whether one prefers the passive or active viewpoint is largely a matter of taste, though it seems more natural to consider physics resulting from invariance under active transformations. The idea is to make this global scaling symmetry a local symmetry by the introduction of a gauge field (the “Weylon”). In the original formulation it was thought that local standards of length would lead to ‘non-integrability of length’, a highly non-desirable feature of Weyl's geometric approach. In the gauge theory approach it is clear that this is not the case. Observable ratios of lengths are integrable and the theory becomes acceptable.

## 4.2 Poincaré-Einstein-Cartan theory

Here we briefly review the gauging arguments of GTG, formulated in the STA. A fuller account can be found in [5].

In field theories the equations of motion are derived from an action principle. For

example the massless Dirac equation arises from minimization of

$$S = \int |d^4x| \langle \nabla \psi \Sigma \gamma_0 \tilde{\psi} \rangle. \quad (4.1)$$

The global symmetry under phase changes can be made local by introducing a covariant derivative involving a gauge field, the vector potential. Adding a term in the gauge-invariant field strength to the action allows one to derive the equations of electromagnetism. This is how the gauge theory of electromagnetism is derived, and we now wish to derive gravity from an analogous gauging argument.

The action has a global displacement symmetry  $x \rightarrow x + c$ , where  $c$  is a constant. However the action is not invariant under local displacements (diffeomorphisms) where each point is moved by a different amount

$$x \rightarrow f(x). \quad (4.2)$$

The first gauge principle of gravity is that the action should be invariant under local displacements of this kind. Clearly this is automatic for scalar and spinor fields if they are undifferentiated since  $\phi'(x) = \phi(x')$ . However the gradient with respect to  $x' = f(x)$  is

$$\nabla_{x'} = \bar{f}^{-1}(\nabla_x, x) \quad (4.3)$$

where we have used the inverse of the adjoint of  $f(a, x) \equiv a \cdot \nabla_x f(x)$ . To ensure that derivatives transform in the same way as the fields we define the displacement gauge function  $\bar{h}(a, x)$  such that

$$\bar{h}'(a, x) = \bar{h}(\bar{f}^{-1}(a, x), f(x)). \quad (4.4)$$

The displacement covariant derivative now transforms as we require

$$\bar{h}(\nabla_x, x) \rightarrow \bar{h}(\nabla_{x'}, x'). \quad (4.5)$$

The  $\bar{h}$ -field is the first gauge field of GTG. From now on we shall drop the explicit  $x$ -dependence, and the local displacement invariant Dirac action is

$$S = \int |d^4x| \det h^{-1} \langle \bar{h}(\nabla) \psi \Sigma \gamma_0 \tilde{\psi} \rangle. \quad (4.6)$$

To see whether the  $\bar{h}$ -field has physical effects consider the displacement covariant derivative

$$L_a = \gamma_a \cdot \bar{h}(\nabla). \quad (4.7)$$

The commutator of these derivatives gives

$$[L_a, L_b] = \underline{e}(L_a \underline{h}_b - L_b \underline{h}_a) \cdot \bar{h}(\nabla) \quad (4.8)$$

where  $\underline{e}_a \equiv \underline{h}_a^{-1}$  is the inverse  $\underline{h}$ -function. Now

$$[L_a \underline{h}_b - L_b \underline{h}_a] \cdot \gamma_c = (\gamma_b \wedge \gamma_a) \cdot (\bar{h}(\nabla) \wedge \bar{h}_c). \quad (4.9)$$

so the condition that the commutator is zero is

$$\bar{h}(\nabla)\wedge\bar{h}_c = 0 \quad \implies \quad \nabla\wedge\bar{e}_c = 0. \quad (4.10)$$

Since this in general will not be the case the  $\bar{h}$  field has physical effects as well as ensuring reparameterization invariance of the action.

Although we have formulated the gauging arguments in the flat Minkowski STA the spacetime we observe will be curved—we observe position gauge invariant quantities, and these will involve the  $\bar{h}$ -function which will not in general be a constant. The metric  $g_{ab}$  which is used extensively in GR can be formed from the  $\bar{h}$ -function and its inverse as

$$g_{ab} = e_a \cdot e_b \quad g^{ab} = \bar{h}^a \cdot \bar{h}^b \quad (4.11)$$

Clearly  $\bar{e}_a$  is related to the vierbein, and this relationship is explained in detail in the next section. It is also useful to define the scalars

$$h_{ab} \equiv \gamma_a \cdot h_b \quad e_{ab} \equiv \gamma_a \cdot e_b \quad (4.12)$$

where naturally  $h_{ab}e^b = \delta_{ac}$ .

The displacement invariant action is also invariant under rotations (active Lorentz transformations) of the fields where

$$\psi \rightarrow R\psi \quad \bar{h}_a \rightarrow R\bar{h}_a\tilde{R} \quad (4.13)$$

and  $R$  is a rotor. The transformation property of the  $\bar{h}$ -field follows from the observation that  $\bar{h}(\nabla)\phi$ , where  $\phi$  is a scalar field, should transform as a covariant vector. We make the rotation invariance local by defining the bivector rotation gauge function  $\Omega_a$ , the ‘spin connection’, that transforms as

$$\Omega_a \rightarrow R\Omega_a\tilde{R} - 2\nabla_a R\tilde{R}. \quad (4.14)$$

This is used to define the rotation-gauge covariant derivative acting on a spinor  $\psi$  or covariant vector  $\mathcal{V}$  (like  $\bar{h}(\nabla)$ ) as

$$D_a\psi \equiv (\nabla_a + \frac{1}{2}\Omega_a)\psi \quad D_a\mathcal{V} \equiv \nabla_a\mathcal{V} + \Omega_a \times \mathcal{V}. \quad (4.15)$$

The explicit form of the derivative in terms of the gauge fields depends on whether it is acting on a spinor or a covariant multivector since they transform differently under rotations. The commutator of the rotation gauge covariant derivatives defines the Riemann tensor  $R_{ab}$ :

$$[D_a, D_b]M \equiv R_{ab} \times M = (\nabla_a\Omega_b - \nabla_b\Omega_a + \Omega_a \times \Omega_b) \times M \quad (4.16)$$

We can now construct a displacement and rotation gauge covariant derivative from  $D_a$  using the  $\bar{h}$ -function

$$\mathcal{D}_a \equiv h^b{}_a D_b. \quad (4.17)$$

Here we start to adopt the useful convention that fully covariant quantities like  $\mathcal{D}_a$  are written in calligraphic type.

Under rotations a covariant tensor  $\mathcal{V}(a)$  will transform as

$$\mathcal{V}(a) \rightarrow R\mathcal{V}(\tilde{R}aR)\tilde{R} \quad (4.18)$$

and the covariant derivative of a tensor is therefore given by

$$\mathcal{D}_a\mathcal{V}_b \equiv L_a\mathcal{V}(\gamma_b) + \omega_a \times \mathcal{V}(\gamma_b) - \mathcal{V}(\omega_a \times \gamma_b). \quad (4.19)$$

where  $\omega_a \equiv h^b{}_a\Omega_b$ . The result is then another covariant tensor of one higher rank, and contractions commute with the covariant derivative. This notation, where the action of the derivative is implied by the transformation properties of what it acts on, is a little different from that in [5] but much closer to traditional tensor notation.

For the action on higher rank tensors we simply extend the definition. For a tensor  $\mathcal{V}_{cd\dots} = \mathcal{V}(\gamma_c, \gamma_d, \dots)$  we have

$$\mathcal{D}_a\mathcal{V}_{cd\dots} \equiv L_a\mathcal{V}(\gamma_c, \gamma_d, \dots) + \omega_a \times \mathcal{V}(\gamma_c, \gamma_d, \dots) - \mathcal{V}(\omega_a \times \gamma_c, \gamma_d, \dots) - \mathcal{V}(\gamma_c, \omega_a \times \gamma_d, \dots) - \dots \quad (4.20)$$

In terms of the scalars  $\mathcal{V}_{bcd\dots} \equiv \gamma_b \cdot \mathcal{V}_{cd\dots}$  and  $\Gamma^c{}_{ab} \equiv \gamma^c \cdot (\omega_a \times \gamma_b) = (\gamma_b \wedge \gamma^c) \cdot \omega_a$  this is

$$\mathcal{D}_a\mathcal{V}_{bcd\dots} = L_a\mathcal{V}_{bcd\dots} - \Gamma^e{}_{ab}\mathcal{V}_{ecd\dots} - \Gamma^e{}_{ac}\mathcal{V}_{bed\dots} - \Gamma^e{}_{ad}\mathcal{V}_{bce\dots} - \dots \quad (4.21)$$

exactly as one would write in tensor notation.

Elements of the background STA like  $\gamma_a$  do not transform under rotations of the fields, and hence  $\mathcal{D}_b\gamma_a = 0$ . The covariant derivative acting on  $\mathcal{A}_a \equiv \gamma_b \cdot \mathcal{A}$ , where  $\mathcal{A}$  is a covariant vector, is therefore

$$\mathcal{D}_a\mathcal{A}_b \equiv \mathcal{D}_a(\gamma_b \cdot \mathcal{A}) \equiv \gamma_b \cdot (\mathcal{D}_a\mathcal{A}) = L_a\mathcal{A}_b - \Gamma^c{}_{ab}\mathcal{A}_c. \quad (4.22)$$

The commutator of the fully covariant derivatives give

$$[\mathcal{D}_a, \mathcal{D}_b]M = \mathcal{R}_{ab} \times M - \mathcal{X}_{abc}\mathcal{D}^cM \quad (4.23)$$

where  $\mathcal{R}_{ab}$  is the covariant Riemann tensor and  $\mathcal{X}_{abc} \equiv (\gamma_b \wedge \gamma_a) \cdot \mathcal{X}_c$  is the torsion. The torsion bivector  $\mathcal{X}_a$  is the natural covariant field strength for the  $\bar{h}$ -function and is given explicitly by

$$\mathcal{X}_a \equiv -e_{ab}\mathcal{D}\wedge\bar{h}^b = \bar{h}(D\wedge\bar{e}^b) \quad (4.24)$$

The action of the covariant derivative here follows from the rotation transformation law of the  $\bar{h}$ -function (which is not a tensor), so

$$\mathcal{D}\wedge\bar{h}^b = \gamma^c \wedge [L_c\bar{h}^b + \omega_c \cdot \bar{h}^b]. \quad (4.25)$$

Similarly, in terms of the inverse we have

$$D\wedge\bar{e}^b = \nabla\wedge\bar{e}^b - \gamma^a \wedge \bar{e}(\Omega_a \cdot \gamma_b) = \nabla\wedge\bar{e}^b + \bar{\Omega}^b{}_c \wedge \bar{e}^c \quad (4.26)$$

where  $\bar{\Omega}_{ab}$  is the vector adjoint of the  $\Omega_c$  function defined by

$$\Omega_c \cdot (\gamma_b \wedge \gamma_a) = \gamma_c \cdot \bar{\Omega}_{ab}. \quad (4.27)$$

The covariant Riemann tensor is the covariant field strength for the rotation gauge field defined by

$$\mathcal{R}_{ab} \equiv h^c{}_a h^d{}_b R_{cd} \quad (4.28)$$

and gives a covariant measure of the curvature. Its contractions define the Ricci tensor and the Ricci scalar

$$\mathcal{R}_b \equiv \gamma^a \cdot \mathcal{R}_{ab} \quad \mathcal{R} \equiv \gamma^a \cdot \mathcal{R}_a. \quad (4.29)$$

The same symbol is used for the Riemann tensor, Ricci tensor and Ricci scalar, with the number of subscripts denoting which is intended (and we shall never need to use  $\mathcal{R}_{ab} = \gamma_a \cdot \mathcal{R}_b$ ). We can also define the adjoint of the Riemann tensor  $\bar{\mathcal{R}}_{ab}$  by

$$(\gamma_a \wedge \gamma_b) \cdot \bar{\mathcal{R}}_{cd} = \mathcal{R}_{ab} \cdot (\gamma_c \wedge \gamma_d). \quad (4.30)$$

To determine the dynamics of the theory we now need to specify an action. We give details of the derivation of the field equations with various actions in later sections. Here we briefly review the results for the Einstein-Hilbert action given by

$$S = \int |d^4x| \det h^{-1} (\mathcal{L}_m - \frac{1}{2\kappa} \mathcal{R}) \quad (4.31)$$

where  $\kappa \equiv 8\pi G$  and  $\mathcal{L}_m$  is the Lagrangian for the other fields present. This gives the Einstein-Cartan theory, which is a subset of the possible Poincaré gauge theories one could obtain using more general actions.

Extremizing with respect to the  $\bar{h}$ -field gives the Einstein equation

$$\mathcal{G}_a \equiv \mathcal{R}_a - \frac{1}{2} \mathcal{R} \gamma_a = \kappa \mathcal{T}_a \quad (4.32)$$

where  $\mathcal{T}_a$  is the stress-energy tensor derived from  $\mathcal{L}_m$

$$\mathcal{T}_a \equiv h^c{}_a \det h \partial_{\bar{h}_c} (\mathcal{L}_m \det h^{-1}). \quad (4.33)$$

Varying with respect to the  $\Omega_a$  field gives a second field equation. In terms of the ‘spin’ of the matter, the bivector  $\mathcal{S}_a \equiv e_{ab} \partial_{\Omega^b} \mathcal{L}_m$ , the equation is

$$\mathcal{X}_a + \mathcal{X} \wedge \gamma_a = -\kappa \mathcal{S}_a. \quad (4.34)$$

where  $\mathcal{X} \equiv \gamma^a \cdot \mathcal{X}_a$  is a vector defined as the contraction of the torsion bivector. When the spin vanishes this ‘structure equation’ determines  $\omega_a$  in terms of  $\bar{h}_a$  in the same way that the connection is determined by the metric in GR. In general (for example for a Dirac Lagrangian) the spin will not vanish and there will be non-vanishing torsion.

The structure equation is algebraic and completely determines the torsion in terms of the spin. The torsion therefore does not propagate — in the vacuum the torsion is zero. This contrasts with the Einstein equation which only determines the Ricci tensor algebraically. The traceless part of the Riemann tensor does not have to be zero in the absence of matter and gives us the long range gravitational force as well as allowing for gravitational waves.

The two field equations we have obtained above in the Geometric Algebra formulation of GTG are equivalent to those obtained for Einstein-Cartan theory using different notation, for example see [33, 35]. Using a more general action, for example including quadratic terms, or terms involving the torsion, leads to more general theories in which the torsion can become propagating. However it is Einstein-Cartan theory that is the simplest extension of General Relativity.



### 4.3 Related formulations: differential forms

In this section we develop a translation scheme to clarify the relation between the GA quantities and those used in differential forms. With the field equations above this demonstrates that the GTG we are using is locally the same as ECKS or  $U_4$  gauge theory. We follow the forms notation and conventions of [32]. If the reader is unfamiliar with forms they can safely skip this section.

Consider a set of four scalar functions  $\{x^\mu(x)\}$  which define a coordinate frame and dual frame

$$e_\mu \equiv \frac{\partial x}{\partial x^\mu} \quad e^\mu \equiv \nabla x^\mu. \quad (4.35)$$

We can define a set of covariant vectors from these via

$$g^\mu \equiv \bar{h}(e^\mu) \quad g_\mu \equiv \underline{e}(e_\mu) \quad (4.36)$$

and metric is then given by

$$g_{\mu\nu} = g_\mu \cdot g_\nu. \quad (4.37)$$

Forms are build up from elements  $dx^\mu$  living in the cotangent space. The exterior derivative translates simply as

$$d \rightarrow \nabla \wedge \quad (4.38)$$

so our translation of the  $dx^\mu$  are just the  $e^\mu$ . However in GA we do not work with elements in the cotangent space, but covariant elements like  $g^\mu$  living in ‘STA-space’. The two spaces are related by the  $\bar{h}$ -function. The general scheme for translation is to convert forms into equivalent GA objects and then apply the  $\bar{h}$ -function to map the result into covariant STA objects.

The forms vierbein generates the metric via

$$g_{\mu\nu} = \eta_{ab} e^a_\mu e^b_\nu \quad (4.39)$$

so the components of the vierbein translate to

$$e^a_\mu \rightarrow g_\mu \cdot \gamma^a. \quad (4.40)$$

The vierbein itself is then

$$e^a = e^a_\mu dx^\mu \rightarrow g_\mu \cdot \gamma^a e^\mu = \bar{e}^a, \quad (4.41)$$

which is of course why we have used this notation for the inverse  $\bar{h}$ -function. There is an unfortunate double use of  $e$  here, but it should be clear from the index whether it is the vierbein or a frame vector.

The components of covariant STA objects are generated by applying  $g_\mu$  and  $g^\mu$  to generate upstairs or downstairs indices as required. For example for a covector  $V$  we have

$$V = V_\mu dx^\mu \rightarrow \mathcal{V} \cdot g_\mu e^\mu = \bar{e}(\mathcal{V}) \quad (4.42)$$

and for the Riemann tensor we have

$$R^\mu{}_{\nu\rho\sigma} = (g^\mu \wedge g_\nu) \cdot \bar{\mathcal{R}}(g_\sigma \wedge g_\rho) \quad (4.43)$$

so that

$$R^a_b = \frac{1}{2}R^a_{b\mu\nu}dx^\mu \wedge dx^\nu \rightarrow \bar{e}(\bar{\mathcal{R}}^a_b) = \bar{R}^a_b. \quad (4.44)$$

Note that the STA Riemann tensor is the adjoint of the forms one. The vierbein is used to convert Latin and Greek indices, for example

$$e^a_\mu A^\mu = g_\mu \cdot \gamma^a \mathcal{A} \cdot g^\mu = \mathcal{A} \cdot \gamma^a = \mathcal{A}^a. \quad (4.45)$$

Similarly the inverse vierbein is given by

$$E_a^\mu = \gamma_a \cdot g^\mu \quad (4.46)$$

and can be used to convert downstairs indices.

The spin-connection  $\omega^a_b$  of differential forms is a 1-form with two antisymmetric indices. This is simply  $\Omega^a_b$ , the adjoint on the  $\Omega_a$ -function used in GTG. Using this we can now translate Cartan's structure equation trivially as follows:

$$de^a + \omega^a_b \wedge e^b = X^a \quad \rightarrow \quad \nabla \wedge \bar{e}^a + \bar{\Omega}^a_b \wedge \bar{e}^b = \bar{e}(\mathcal{X}^a) \quad (4.47)$$

$$\implies \bar{h}(D \wedge \bar{e}^a) = \mathcal{X}^a \quad (4.48)$$

where  $\mathcal{X}_a$  defines the covariant torsion bivector (which vanishes in GR). Similarly the Riemann tensor is given by

$$d\omega^a_b + \omega^a_c \wedge \omega^c_b \quad \rightarrow \quad \dot{\nabla} \wedge \gamma^c (\gamma_b \wedge \gamma^a) \cdot \dot{\Omega}_c + (\gamma_c \wedge \gamma^a) \cdot \dot{\Omega}_e (\gamma_b \wedge \gamma^c) \cdot \dot{\Omega}_f \gamma^e \wedge \gamma^f \quad (4.49)$$

$$= -\frac{1}{2}\gamma^e \wedge \gamma^c (\gamma_b \wedge \gamma^a) \cdot (\nabla_e \Omega_c - \nabla_c \Omega_e + \Omega_e \times \Omega_c) \quad (4.50)$$

$$= \bar{R}^a_b = \bar{e}(\bar{\mathcal{R}}^a_b) \quad (4.51)$$

The GTG formulation is therefore easily related to similar gauge theory formulations with differential forms. However using GA we have the advantage of being able to use the geometric and dot products, which can make life a little easier. We also have no need to worry about tangent and cotangent spaces, and can construct manifestly covariant objects in the STA. The main advantage of GTG in the formulation we are using is that equations can be written in terms of covariant objects that are directly observable quantities.

As we have seen the GA objects are the 'inside out' versions of the forms ones—the natural GA fields are the adjoints of the forms ones. This is because in the GA formulation we have considered gravitational fields defined on a flat background space, and then gauged rotations of the fields. In the forms formulations the background space is curved and one gauges the rotations of a local orthonormal frame. The GA version is therefore rather more restrictive as we have assumed a trivial background topology. However locally the theories are equivalent, and if needed one could no doubt modify the background STA to have a non-trivial topological structure.

One can form the inner product of forms using the Hodge dual. In the GA approach we have no need to do this since we can just take the inner product of covariant vectors. However we can nonetheless write down a translation. For two cotangent vectors  $A$  and  $B$  the inner product is defined as

$$A \wedge^* B = g^{\mu\nu} A_\mu B_\nu |g|^{\frac{1}{2}} dx^1 \wedge \dots \wedge dx^n. \quad (4.52)$$

Now

$$\mathcal{A} \cdot \mathcal{B} = A^\mu B^\nu g_\mu \cdot g_\nu = g_{\mu\nu} A^\mu B^\nu \quad (4.53)$$

so defining  $C = {}^*B$  the forms definition gives to

$$\bar{e}(\mathcal{A}) \wedge C = I \mathcal{A} \cdot \mathcal{B} \det h^{-1} \quad (4.54)$$

$$\implies \mathcal{A} \wedge C = I \mathcal{A} \cdot \mathcal{B} \quad (4.55)$$

$$\implies C = -I \mathcal{B}. \quad (4.56)$$

$$(4.57)$$

At the level of covariant objects the dual therefore simply corresponds to multiplication by  $I$ .

## 4.4 General Relativity

In this section we review some of the main features of GR, expressed in the language of GTG when the spin vanishes and the theories are locally equivalent.

In the absence of spin the torsion  $\mathcal{X}_a$  is zero and the commutator of the covariant derivatives gives the Ricci ‘identity’ of GR

$$[\mathcal{D}_a, \mathcal{D}_b]M = \mathcal{R}_{ab} \times M. \quad (4.58)$$

Note that in GTG this is really an equation, it is only true if the torsion vanishes.

Since the torsion vanishes we have  $\mathcal{D} \wedge \bar{h}_a = 0$ . Using the Ricci identity it follows that

$$\mathcal{D} \wedge \mathcal{D} \wedge \bar{h}_a = 0 \quad \implies \quad \bar{\mathcal{R}}_{ab} \wedge \gamma^b = 0. \quad (4.59)$$

This is the ‘cyclic identity’ of GR, and summarizes the symmetries of the Riemann tensor. In particular we have

$$\gamma_c \cdot (\bar{\mathcal{R}}_{ab} \wedge \gamma^b) = 0 \quad \implies \quad \bar{\mathcal{R}}_{ac} - \gamma^b \wedge (\gamma_c \cdot \bar{\mathcal{R}}_{ab}) = 0 \quad (4.60)$$

$$\implies \quad \bar{\mathcal{R}}_{ac} + \gamma^b \wedge (\gamma_a \cdot \mathcal{R}_{cb}) = 0 \quad (4.61)$$

$$\implies \quad \bar{\mathcal{R}}_{ac} + \mathcal{R}_{ca} - \gamma_a \cdot (\gamma^b \wedge \mathcal{R}_{cb}) = 0 \quad (4.62)$$

$$\implies \quad \bar{\mathcal{R}}_{ab} = \mathcal{R}_{ab} \quad (4.63)$$

so the Riemann tensor is symmetric. The Riemann tensor is antisymmetric on  $a, b$ , and so can be viewed as a bivector valued function of a bivector. The symmetry relation above gives 16 constraints and the Riemann tensor therefore has a total of  $6 \times 6 - 16 = 20$  degrees of freedom.

The Weyl tensor  $\mathcal{W}_{ab}$  is defined as the symmetric traceless part of the Riemann tensor so that

$$\gamma^a \mathcal{W}_{ab} = 0. \quad (4.64)$$

(We give the general decomposition of the Riemann tensor in a later chapter.) The Weyl tensor also has the property that [5]

$$\mathcal{W}_{ab}^* = {}^* \mathcal{W}_{ab} \equiv I \mathcal{W}_{ab} \quad (4.65)$$

where the ‘interior’ dual is given by

$$\mathcal{W}_{ab}^* = \frac{1}{2} I \gamma_a \wedge \gamma_b \wedge \gamma^d \wedge \gamma^c \mathcal{W}_{cd} = \mathcal{W}(I \gamma_a \wedge \gamma_b). \quad (4.66)$$

The Bianchi identity follows from the Jacobi identity for the covariant derivatives

$$[\mathcal{D}_a, [\mathcal{D}_b, \mathcal{D}_c]] + \text{cyclic permutations} = 0. \quad (4.67)$$

Using the Ricci identity to replace the commutators of the covariant derivatives this can be written

$$\gamma^a \wedge \gamma^b \wedge \gamma^c \mathcal{D}_a \mathcal{R}_{bc} = 0. \quad (4.68)$$

Using the symmetry of the Riemann tensor this then gives the Bianchi identity

$$\mathcal{D} \wedge \mathcal{R}_{ab} = 0. \quad (4.69)$$

Contracting with  $\gamma_a \wedge \gamma_b$  gives the contracted Bianchi identity

$$\mathcal{D} \cdot \underline{\mathcal{G}}_a = 0 \quad (4.70)$$

which, from the Einstein equation, implies conservation of stress-energy,  $\mathcal{D} \cdot \underline{\mathcal{T}}_a = 0$ . Since we have used the Ricci equation these Bianchi ‘identities’ will only hold in the absence of torsion.

The GTG equivalent of the GR infinitesimal diffeomorphism gauge transformation law is readily obtained. Consider an infinitesimal displacement to

$$x' = f(x) = x + \delta\eta \quad (4.71)$$

The inverse of the  $\bar{h}$ -function transforms as

$$\underline{e}(a, x) \rightarrow \underline{e}(f(a), x') \quad (4.72)$$

where here

$$\underline{f}(a) = a + \nabla_a \delta\eta. \quad (4.73)$$

Substituting for  $x'$  the transformation law for  $\underline{e}_a$  becomes (to first order)

$$\underline{e}_a \rightarrow \underline{e}_a + \delta\eta \cdot \nabla \underline{e}_a + \underline{e}(\nabla_a \delta\eta). \quad (4.74)$$

A Killing vector is usually defined by a symmetry of the metric. Here a Killing vector  $\mathcal{K}$  generates an isometry when  $\eta = \bar{h}(\mathcal{K})$  and the change in  $\underline{e}_a$  can be removed by a rotation gauge transformation. To determine  $\mathcal{K}$  we therefore want the rotation gauge covariant information in

$$\mathcal{K} \cdot \bar{h}(\nabla) \underline{e}_a + \underline{e}(\nabla_a \bar{h}(\mathcal{K})) = 0 \quad (4.75)$$

which can be written

$$L_a \mathcal{K} + \mathcal{K}^b \underline{e}(L_a \bar{h}_b - L_b \bar{h}_a) = 0. \quad (4.76)$$

Assuming zero torsion and using the structure equation this becomes

$$\mathcal{D}_a \mathcal{K} + \gamma_a \cdot \omega_b \mathcal{K}^b = 0. \quad (4.77)$$

To extract covariant information we dot with  $\gamma_c$  and symmetrize giving Killing's equation

$$\mathcal{D}_{(a}\mathcal{K}_{c)} = 0. \quad (4.78)$$

We could equally well have derived this by insisting the metric tensor (which is rotation invariant) were displacement invariant. However it is useful to see how the same information can be extracted directly from the  $\bar{h}$ -function.

## 4.5 Extending the gauge symmetries

So far we have only considered Poincaré gauge theory, in which displacements and local rotations of the fields are gauged. However we could consider enlarging the symmetries. The most general theories usually considered are metric-affine theories [45] in which a general linear transformation of the fields is gauged. These include new scale and shear transformations, as well as the rotations, and can in general be written

$$\mathcal{V}_a \rightarrow L_{ab}\mathcal{V}^b. \quad (4.79)$$

A shear transformation is governed by the symmetric traceless part of  $L_{ab}$  whilst the trace determines the scale transformation. The anti-symmetric part gives the rotation part we already have in the Poincaré gauge theory.

The shear transformation can be written in terms of two orthogonal vectors  $A$  and  $B$  as  $L_{ab} = A_{(a}B_{b)}$ . However there are immediate difficulties in writing down a shear transformation for a spinor. A spin vector  $\psi\gamma\tilde{\psi}$  should transform like a vector, but one cannot devise a transformation of  $\psi$  that would effect this. Metric-affine theories including shears have to resort to infinite-dimensional spinor representations in order to make things work [45].

The scale part of the transformation however poses no such problems, and in the remainder of this chapter we consider Weyl-Cartan theory, Poincaré gauge theory with the addition of gauged local scale invariance of the fields. Scale invariance is well motivated: the renormalizable field theories of the standard model are scale invariant prior to spontaneous symmetry breaking, and one of the fundamental problems with GR is that it is non-renormalizable because it is not scale invariant.

The only other obvious transformation to consider is where a spinor transforms as  $\psi \rightarrow e^{I\beta}\psi$ . This leaves invariant terms in the electroweak Lagrangian. However since  $e^{I\beta}$  commutes with  $\psi$  this is most readily understood as an 'internal' symmetry representing the  $U(1)$  electroweak symmetry. (The  $SU(3)$  symmetry can be considered as symmetry under 'internal rotations' of the spinor).

## 4.6 Weyl-Cartan gravity

We now extend GTG as formulated by Lasenby *et al.* [5] to include local scale transformations. We define a scale transformation for objects with dimensions of mass as  $M \rightarrow e^\lambda M$ . Length has dimensions of  $(\text{mass})^{-1}$  and so objects with dimensions of length transform

as  $L \rightarrow e^{-\lambda}L$ . In general a quantity  $M$  is described as having conformal weight  $n$  if it transforms as

$$M \rightarrow e^{n\lambda}M. \quad (4.80)$$

The  $\bar{h}$ -field has weight one and so transforms as

$$\bar{h}_a \rightarrow e^\lambda \bar{h}_a. \quad (4.81)$$

We now try to construct an action that is locally invariant under this transformation (i.e. when  $\lambda$  can be a function of position).

The Lagrangian for a massive spinor has a term  $m\psi\tilde{\psi}$ . For this to have dimensions of  $(length)^{-4}$  to yield a dimensionless action the spinor field must have dimensions of  $(mass)^{3/2}$ . According to our definition spinors therefore have weight  $3/2$ . The kinetic term in the Dirac Lagrangian is then automatically locally scale invariant as can readily be checked.

However when we have derivatives of a scalar field the kinetic term is not automatically scale invariant and we need to define a new gauge field in order to maintain local invariance. The new covariant derivative is given by

$$\check{D}_a \equiv D_a + nK_a \quad (4.82)$$

where under the scaling the new vector gauge field  $K$  transforms as

$$K \rightarrow K - \nabla\lambda. \quad (4.83)$$

The new scale covariant derivative satisfies Leibnitz rule and scales as

$$\check{D}_a M \rightarrow e^{n\lambda} \check{D}_a M. \quad (4.84)$$

The fully covariant derivative is formed in the usual way

$$\check{\mathcal{D}}_a = h^b{}_a \check{D}_b \quad (4.85)$$

and the covariant version of the gauge field is therefore given by  $\mathcal{K} \equiv \bar{h}^a K_a$ .

The transformation law for the  $K$  field is the same as that for the photon, which was what lead Weyl to an identification. Here we regard  $K$  as a new field associated with new physics since we now know that it is not scale symmetry that generates electromagnetism.

In GR the Ricci scalar does not transform homogeneously and can be combined with derivatives of a scalar field to produce a combination which is overall scale covariant:

$$\phi^2 \mathcal{R} + 6\partial_\mu \phi \partial^\mu \phi. \quad (4.86)$$

In GTG the Riemann tensor transforms homogeneously at the level of the action and the analogous term would not be scale covariant. However in GTG we have the torsion tensor that does not transform homogeneously. Indeed under a scaling the contraction of the torsion tensor transforms like  $-3\mathcal{K}$ :

$$\mathcal{X} \rightarrow e^\lambda [\mathcal{X} + 3\mathcal{D}\lambda]. \quad (4.87)$$

One could therefore construct the combination

$$(\mathcal{D}\phi - \frac{1}{3}\mathcal{X}\phi) \cdot (\mathcal{D}\phi - \frac{1}{3}\mathcal{X}\phi) \quad (4.88)$$

that transforms homogeneously, and could form a valid term in the action integral. More generally we could replace  $\mathcal{K}$  with  $-\frac{1}{3}\mathcal{X}$  everywhere to generate scale invariant actions without introducing a new  $K$  field at all. For the moment however we shall stick with the gauging argument and treat  $K$  as an independent field, at least at the level of the action. This is the procedure most consistent with the gauging arguments for the other symmetries.

The commutator of the fully covariant derivatives now has an additional term for the field strength  $\mathcal{H}$  of the  $K$  field,

$$[\check{\mathcal{D}}_a, \check{\mathcal{D}}_b]M = \check{\mathcal{R}}_{ab} \times M - \check{\mathcal{X}}_{abc} \check{\mathcal{D}}^c M + n\mathcal{H}_{ab}M. \quad (4.89)$$

The field strength is given explicitly by

$$\mathcal{H} \equiv \bar{h}(\nabla \wedge K) \quad \mathcal{H}_{ab} \equiv (\gamma_b \wedge \gamma_a) \cdot \mathcal{H} = \underline{h}_b \wedge \underline{h}_a \cdot (\nabla \wedge K) \quad (4.90)$$

and the scale covariant torsion is defined in terms of the scale covariant derivative

$$\check{\mathcal{X}}_a \equiv -e_{ab} \check{\mathcal{D}} \wedge \bar{h}^b = -e_{ab} \gamma^c \wedge [L_c \bar{h}^b + \omega_c \cdot \bar{h}^b + \mathcal{K}_c \bar{h}^b]. \quad (4.91)$$

We now proceed to construct a gauge invariant action integral of the form

$$S = \int |d^4x| \det h^{-1} \mathcal{L} \quad (4.92)$$

where  $\mathcal{L}$  must have weight four. As for electromagnetism we can form a kinetic term in the  $K$ -field is proportional to  $\mathcal{H} \cdot \mathcal{H}$ . Since  $\det h$  and  $\mathcal{H} \cdot \mathcal{H}$  have weight four this term is scale invariant. However the Ricci scalar has conformal weight two, so this cannot be in the Lagrangian on its own. We therefore follow Dirac [36] by introducing a scalar field  $\phi$  of conformal weight one and coupling its square to the Ricci scalar. The total Lagrangian is now

$$\mathcal{L} = \frac{\nu^2}{2} \check{\mathcal{D}}\phi \cdot \check{\mathcal{D}}\phi - \frac{\alpha}{2} \phi^2 \mathcal{R} - \Lambda \phi^4 + \frac{\beta^2}{2} \mathcal{H} \cdot \mathcal{H} + \mathcal{L}_m \quad (4.93)$$

where  $\mathcal{L}_m$  is the Lagrangian of any other fields. There are other terms quadratic in the Riemann tensor that could be introduced, but we defer a discussion of these to the next chapter. To determine the classical field equations we now vary the action with respect to the four independent fields  $\bar{h}_a$ ,  $\underline{\Omega}_a$ ,  $K$  and  $\phi$ . This procedure is consistent with the gauge theory formulation and leads to results which differ somewhat from those obtained within extensions of GR.

## 4.7 The field equations

The field equations are derived by using the multivector function form of the Euler-Lagrange equation:

$$\partial_{M_a}(\mathcal{L} \det h^{-1}) = \nabla^b(\partial_{M_{a,b}} \mathcal{L} \det h^{-1}) \quad (4.94)$$

where as usual  $M_{a,b}$  denotes  $\nabla_b M_a$ .

Since  $\bar{h}_a$  is undifferentiated in the action the  $\bar{h}_a$  equation is

$$\begin{aligned} \partial_{\bar{h}_a} \mathcal{L} &= 0 \\ \implies \alpha \phi^2 \underline{\mathcal{G}}_a &= \underline{\mathcal{T}}_a \end{aligned} \quad (4.95)$$

where  $\underline{\mathcal{G}}_a$  is the Einstein tensor defined by

$$\underline{\mathcal{G}}_a = \underline{\mathcal{R}}_a - \frac{1}{2} a \mathcal{R} \quad (4.96)$$

and  $\underline{\mathcal{T}}_a$  is the stress-energy tensor defined by

$$\underline{\mathcal{T}}_a = h^b{}_a \det h \partial_{\bar{h}_b} ([\mathcal{L} + \frac{1}{2} \alpha \phi^2 \mathcal{R}] \det h^{-1}). \quad (4.97)$$

This is the new modified Einstein equation. Similarly variation with respect to  $\Omega_a$  gives the modified structure equation

$$(\check{\mathcal{D}}\phi^2) \wedge \gamma_a - \phi^2 (\check{\mathcal{X}}_a + \check{\mathcal{X}} \wedge \gamma_a) = \alpha^{-1} \underline{\mathcal{S}}_a \quad (4.98)$$

where  $\underline{\mathcal{S}}_a$  is the ‘spin’ of the matter as before. In the case where the spin is of Dirac type the contraction of the spin tensor vanishes [5],  $\gamma^a \cdot \underline{\mathcal{S}}_a = 0$ , so that

$$\gamma^a \cdot [(\check{\mathcal{D}}\phi^2) \wedge \gamma_a - \phi^2 (\check{\mathcal{X}}_a + \check{\mathcal{X}} \wedge \gamma_a)] = 0 \quad (4.99)$$

$$\implies 2\phi^2 \check{\mathcal{X}} - 3\check{\mathcal{D}}\phi^2 = 0. \quad (4.100)$$

We shall assume this is the case, and can therefore write the field equation as

$$\phi^2 (\check{\mathcal{X}}_a + \frac{1}{3} \check{\mathcal{X}} \wedge \gamma_a) = -\alpha^{-1} \underline{\mathcal{S}}_a. \quad (4.101)$$

Variation with respect to  $K$  gives

$$\nu^2 \underline{h}(\phi \check{\mathcal{D}}\phi) + \partial_K \mathcal{L}_m = \frac{1}{2} \det h \beta^2 \nabla^a (\partial_{K,a} [(\nabla \wedge K) \cdot \underline{h}(\mathcal{H})] \det h^{-1}) \quad (4.102)$$

$$= -\beta^2 \det h \nabla \cdot [\underline{h}(\mathcal{H}) \det h^{-1}]. \quad (4.103)$$

Rearranging this becomes

$$\nu^2 \phi \check{\mathcal{D}}\phi = -\beta^2 \underline{e}(\nabla \cdot G) \det h - \underline{e}(\partial_K \mathcal{L}_m) \quad (4.104)$$

where  $G = \underline{h}(\mathcal{H}) \det h^{-1}$ . Finally the  $\phi$  equation gives

$$-4\phi^3 \Lambda - \alpha \phi \mathcal{R} + \partial_\phi \mathcal{L}_m = -\nu^2 \mathcal{K} \cdot \check{\mathcal{D}}\phi + \nu^2 \det h \nabla^a [\partial_{\phi,a} (\det h^{-1} \phi_b \bar{h}^b) \cdot \check{\mathcal{D}}\phi] \quad (4.105)$$

$$= -\nu^2 \mathcal{K} \cdot \check{\mathcal{D}}\phi + \nu^2 (\check{\mathcal{X}} - \Omega_a \times \bar{h}^a + 3\mathcal{K} + \bar{h}(\nabla)) \cdot \check{\mathcal{D}}\phi \quad (4.106)$$

$$= \nu^2 (\check{\mathcal{D}} + \check{\mathcal{X}}) \cdot \check{\mathcal{D}}\phi \quad (4.107)$$

Using (4.100) this becomes

$$4\Lambda + \alpha \phi^{-2} \mathcal{R} - \phi^{-3} \partial_\phi \mathcal{L}_m = -\nu^2 \check{\mathcal{D}} \cdot (\phi^{-3} \check{\mathcal{D}}\phi). \quad (4.108)$$



and we can use (4.95) to substitute for  $\mathcal{R}$  giving

$$-4\Lambda + \phi^{-4}\mathcal{T} + \phi^{-3}\partial_\phi\mathcal{L}_m = \nu^2\check{\mathcal{D}} \cdot (\phi^{-3}\check{\mathcal{D}}\phi). \quad (4.109)$$

where  $\mathcal{T} = \gamma^a \cdot \underline{\mathcal{T}}_a$ .

In [5] there was considerable analysis of the quantity  $\mathcal{X}$  defined by the contraction of the torsion tensor. In the absence of the scale symmetry the field equations implied that it vanished if the spin tensor had zero contraction. Here the Dirac Lagrangian is automatically locally scale invariant. It gives rise to the field equation

$$\check{\mathcal{D}}\psi\Sigma = m\psi\gamma_0 + \frac{1}{2}\check{\mathcal{X}}\psi\Sigma \quad (4.110)$$

where the  $K$  in the derivative and in  $\check{\mathcal{X}}$  can be cancelled from both sides to give

$$\mathcal{D}\psi\Sigma = m\psi\gamma_0 + \frac{1}{2}\mathcal{X}\psi\Sigma. \quad (4.111)$$

Instead of the minimally coupled equation

$$\check{\mathcal{D}}\psi\Sigma = m\psi\gamma_0 \quad (4.112)$$

we find that the local scale invariance is maintained by the contraction of the torsion tensor which, as we know, transforms like the gauge field under scalings. Interestingly we cannot find any action that would give rise to the latter equation.

## 4.8 The Einstein gauge

In a scale gauge theory observables must be scale gauge invariant. In particular we can use the scalar field to construct scale-invariant quantities from the non-scale invariant fields

$$\bar{h}'_a = \bar{h}_a/\phi \quad \text{and} \quad \underline{\mathcal{S}}'_a = \underline{\mathcal{S}}_a/\phi^3 \quad (4.113)$$

as long as  $\phi$  is non-zero. The torsion is then given in terms of the torsion of  $\bar{h}'_a$  by

$$\check{\mathcal{X}}'_a = \phi\check{\mathcal{X}}'_a - (\check{\mathcal{D}}'\phi)\wedge\gamma_a. \quad (4.114)$$

Using these relations the modified structure equation, equation (4.98), becomes

$$\mathcal{X}'_a + \mathcal{X}' \wedge \gamma_a = -\kappa \underline{\mathcal{S}}'_a \quad (4.115)$$

where  $\kappa = 1/\alpha$ . If  $\alpha = 1/8\pi G$  this is the structure equation of GTG. Similarly we can recover Einstein's equation by using the invariant variables

$$\underline{\mathcal{G}}'_a = \phi^{-2}\underline{\mathcal{G}}_a \quad \text{and} \quad \underline{\mathcal{T}}'_a = \phi^{-4}\underline{\mathcal{T}}_a \quad (4.116)$$

in equation (4.95). We can set the dashed variables equal to the undashed variables by fixing the scale gauge so that  $\phi = 1$ , called the Einstein gauge since it reduces the  $\bar{h}$  equation to Einstein's equation. This is possible in the presence of stress-energy since  $\phi$  must be

non-zero and the Lagrangian is symmetric under  $\phi \rightarrow -\phi$ . However a more general gauge would be needed to discuss the possibility of  $\phi = 0$  and at the quantum level we anticipate other gauges being more useful. With  $\phi = 1$  we see that the quartic  $\phi$  term in the action would give rise to a cosmological constant if  $\Lambda \neq 0$  (see [47] for a discussion).

Using (4.101) and fixing  $\phi = 1$  in the  $K$ -equation (4.104) we have

$$\nu^2 \mathcal{K} + \underline{e}(\partial_K \mathcal{L}_m) = -\beta^2 \underline{e}(\nabla \cdot G) \det \underline{h} \quad (4.117)$$

$$= -\beta^2 [\mathcal{D} \cdot \mathcal{H} - \mathcal{S}^a \cdot (\gamma_a \wedge \mathcal{H})]. \quad (4.118)$$

When the spin is a trivector (as for electroweak spinors) the spin  $\mathcal{S}_a = \mathcal{S} \gamma_a$  where  $\mathcal{S} = \psi I \gamma \tilde{\psi}$  and  $\gamma$  is a basis vector. In this case we can write

$$\underline{e}(\partial_K \mathcal{L}_m) + \nu^2 \mathcal{K} = -\beta^2 (\mathcal{D} \cdot \mathcal{H} - \kappa \mathcal{S} \cdot \mathcal{H}). \quad (4.119)$$

Taking the divergence and making use of the relation [5]

$$\mathcal{D} \wedge [\bar{h}(A_r)] = \bar{h}(\nabla \wedge A_r) + \kappa \langle \mathcal{S} \bar{h}(A_r) \rangle_{r+1} \quad (4.120)$$

we get

$$\nu^2 \mathcal{D} \cdot \mathcal{K} + \mathcal{D} \cdot \underline{e}(\partial_K \mathcal{L}_m) = -\beta^2 (I \bar{h}(\nabla \wedge [\bar{e}(I \mathcal{D} \cdot \mathcal{H})]) - \kappa \mathcal{D} \cdot (\mathcal{S} \cdot \mathcal{H})). \quad (4.121)$$

Now

$$I \mathcal{D} \cdot \mathcal{H} = -\bar{h}(\nabla \wedge [\bar{e}(I \mathcal{H})]) + \kappa I \mathcal{S} \cdot \mathcal{H} \quad (4.122)$$

so that

$$I \bar{h}(\nabla \wedge [\bar{e}(I \mathcal{D} \cdot \mathcal{H})]) = \kappa \mathcal{D} \cdot (\mathcal{S} \cdot \mathcal{H}) \quad (4.123)$$

and finally

$$\nu^2 \mathcal{D} \cdot \mathcal{K} = -\mathcal{D} \cdot [\underline{e}(\partial_K \mathcal{L}_m)]. \quad (4.124)$$

Since  $K$  can only enter the matter Lagrangian via derivatives of scalar fields we see that  $K$  is divergenceless in the absence of other scalar fields. To see what sort of field  $K$  is we can write equation (4.119) as

$$\mathcal{D}^2 \mathcal{K} - \mathcal{D}(\mathcal{D} \cdot \mathcal{K}) - \mathcal{K}_a \bar{\mathcal{R}}^a - \kappa(\mathcal{D} - \kappa \mathcal{S}) \cdot (\mathcal{K} \cdot \mathcal{S}) + m^2 \mathcal{K} = 0 \quad (4.125)$$

where  $m = \nu/\beta$ . When  $\mathcal{L}_m$  is independent of  $K$  and in the absence of torsion this is the equation of a massive vector field of mass  $m$ :

$$\mathcal{D}^2 \mathcal{K} - \mathcal{K}_a \bar{\mathcal{R}}^a + m^2 \mathcal{K} = 0 \quad (4.126)$$

Since  $K$  only couples to scalar fields directly it is essentially non-interacting and therefore makes a potential dark matter candidate [38]. For spherically symmetric solutions for a massive vector field and further references see [48].

The mass here differs from that found in Weyl's geometry where the only gravitational field in the action is the metric. This difference in masses is typical of the result obtained when varying one rather than two fields in the action [49]. In GR this is equivalent to the

difference between the Palatini principle, where the connection is varied as an independent field, versus variation with respect to the metric where the connection is predetermined. In the gauge theory approach it is clear that  $\Omega_a$  should be varied independently of  $\bar{h}_a$ . The difference in predicted masses is probably trivial since the theory does not predict the value of  $\beta$  (unless perhaps  $\beta = 1$ ). However the difference is much more radical with a more complicated (e.g. quadratic) Lagrangian as we shall see in the next chapter.

Fixing  $\phi = 1$  and using Einstein's equation the  $\phi$  equation (4.109) becomes

$$\nu^2 \mathcal{D} \cdot \mathcal{K} - \partial_\phi \mathcal{L}_m|_{\phi=1} - \mathcal{T}^{\mathcal{L}_m} = 0. \quad (4.127)$$

Substituting for  $\mathcal{D} \cdot \mathcal{K}$  from (4.124) we have the equation

$$\partial_\phi \mathcal{L}_m|_{\phi=1} + \mathcal{T}^{\mathcal{L}_m} + \mathcal{D} \cdot [e(\partial_K \mathcal{L}_m)] = 0 \quad (4.128)$$

that must be satisfied by any locally scale invariant matter Lagrangian. We note that the trace of the stress-energy tensor is not zero, as expected for a realistic theory [47].

In summary, we have recovered GTG spin-torsion theory but with an additional massive vector field that does not couple to spinors. This is very similar to conformal extensions of GR. However the gauge theory  $\Omega_a$  field is weight zero while the connection in GR contains derivatives of the metric tensor and must be modified to contain  $K$ s. This is the key difference between the gauge theory formulation and the introduction of local scale invariance in GR via a new gauge field. It explains why we get different results from the field equations, though it is not clear that there any observable differences. The gauge theory approach would seem to be more consistent.

## 4.9 Discussion

Tsamis and Woodward [47] have shown that some scalar-metric theories of GR are identical to their pure metric analogues. The theories they analysed were generated by forming scale invariant combinations of the Ricci scalar and derivatives of the scalar field, and the action is

$$\bar{S}[\phi, g, \psi_i] = S[\phi^2 g, \phi^{-n} \psi_i] \quad (4.129)$$

where  $\psi_i$  are fields with weight  $n$  (we have defined weights to be the negative of what they are usually defined to be). Similarly they show that the quantum theory in the Einstein gauge is identical to the pure metric theory. They conclude that local scale invariance is a sham symmetry without dynamical import.

In our approach much of their argument goes through. For instance relations (4.113) show that  $\phi$  is still really just a scale factor for the other fields. However since we have to introduce a new gauge field there does seem to be some new physics. The bare theory we end up with is identical to the non-scale invariant theory with the addition of a minimally coupled massive vector field. The  $\phi$  field is still without dynamical import, but its kinetic term cannot be generated directly from the non-scale invariant Lagrangian by substituting  $\bar{h}_a = \phi^{-1} \bar{h}_a$  as it could be in GR. We now have the  $K$  field which transforms inhomogeneously so all we can say is that

$$\bar{S}[\phi, \bar{h}_a, \Omega_a, \psi_i, K] = S[\phi^{-1} \bar{h}_a, \Omega_a, \phi^{-n} \psi_i, K + \phi^{-1} \nabla \phi], \quad (4.130)$$

which is easily proved by performing a scaling proportional to  $\phi^{-1}$ . Similar remarks of course apply to theories in GR where a new gauge field is introduced.

If we had used the contraction of the torsion tensor to construct a scale invariant action integral then the argument would have had full force and the theory would be equivalent to a non-scale invariant theory with a torsion squared term in the action. Using a general vector gauge field instead we have the possibility of a new physical field, as well being consistent with the gauging arguments for the other symmetries.

In order to generate a more physically interesting theory it is possible to consider whether  $\phi$  might have some internal structure. A tempting possibility is identifying  $\phi$  with the modulus of the Higgs field. The standard model Lagrangian density can be written

$$\mathcal{L} = \langle -\mathcal{D}\psi I\gamma_0\tilde{\psi} + \mathcal{D}\chi I\gamma_0\tilde{\chi} - \frac{1}{4}\mathcal{F}_{ab}\cdot\mathcal{F}^{ab} + \frac{1}{2}\mathcal{I}\cdot\mathcal{I} + \nu^2(\mathcal{D}^a\phi_H)^\dagger\mathcal{D}_a\phi_H - 2m_e\psi\phi_H\tilde{\chi} \rangle - V \quad (4.131)$$

which is scale covariant if  $V$  is. (The details of our translation into geometric algebra are not important here). To fit in with conformal gravity the electroweak Lagrangian must be locally scale invariant. This means that the covariant derivatives must be replaced with the scale covariant invariant ones and that  $V$  must be of weight four. There are two possibilities. We could have

$$V = \eta(\phi_H^2 - \nu_H^2\phi^2)^2 \quad (4.132)$$

to recover the standard model when we use the Einstein gauge. Alternatively we could identify the modulus of the Higgs field with  $\phi$  in which case we must have  $V = \Lambda\phi^4$ . In this case there is no symmetry breaking; instead the symmetry remains unbroken and the observed particle spectrum comes from working in the Einstein gauge. Of course observables must be gauge invariant so this provides an interesting alternative to the usual Higgs mechanism. The main prediction is that the Higgs particle does not exist since there is now one extra degree of freedom — the scale gauge freedom. The  $\phi$  field can now be set exactly to one. This also means that  $K$  does not couple to any standard model fields except via gravity.

This Higgs free model has been developed before [41, 42]. If we do not adopt the Higgs free model we gain some terms in the standard model from the coupling of  $K$  to the Higgs field. This has the advantage of making the  $K$  field potentially detectable, and allows standard model physics to remain otherwise unchanged (though we have not checked the quantum implications).

## 4.10 Conclusions

We have reviewed the formulation of GTG within Geometric Algebra and shown how it relates to formulations in terms of differential forms. The theories are essentially equivalent, and in the absence of torsion GTG is locally equivalent to GR.

We then showed that it is possible to develop a locally scale invariant GTG. In the absence of other scalar fields the theory is equivalent to adding a massive vector field to the non-scale invariant theory. It is therefore consistent with observational tests of GR. The only physical interest seems to come from considering the new field as a potential dark matter candidate.

With the addition of other scalar fields things may be more interesting. It is possible to identify the  $\phi$  field with the modulus of the Higgs field, giving the bold prediction that the Higgs particle does not exist. There is a good possibility that this may be discounted shortly by the discovery of a Higgs, in which case physical interest enters through the coupling of the  $K$  field to the Higgs. This could potentially give a mechanism for generating dark-matter  $K$ s in the early universe. A cosmological constant terms also arises naturally. However the theory makes no predictions for the magnitude of these terms.

The scale invariant formalism may also have mathematical uses. The possibility of not using the Einstein gauge gives a different mathematical approach to essentially the same physical problem. One could consider fixing the scale gauge by fixing  $K$  rather than  $\phi$ , which could have applications in constructing a quantum theory.



## Chapter 5

# Quadratic Lagrangians and Topology

In this chapter we analyse some topological configurations of the spin connection field in Poincaré Gauge Theory Gravity. There are two topological invariants that we show arise from the scalar and pseudoscalar parts of a single integral. We explore the link with Yang-Mills instantons in Euclidean gravity. Neither of the topological action integrals contribute to the classical field equations and they therefore restrict the number of independent quadratic terms we can have in the GTG Lagrangian. There are ten independent terms quadratic in the Riemann tensor and the topological invariants reduce these to eight possible classically independent terms. The resulting field equations for the parity non-violating terms are presented. The motivation for studying quadratic Lagrangians stems in part from the analysis of scale invariant theories in the previous chapter, and we also expect quadratic term to be essential to the formulation of a sensible quantum theory.

### 5.1 Introduction

It is possible to construct quadratic terms for the gravitational Lagrangian that are total divergences. In the gauge theory approach to gravity these are topological invariants and are given by boundary terms in the action integral. They do not affect the classical field equations, though they could become important in a quantum theory. The invariants have a natural analogue in Euclidean gravity as the winding numbers of Yang-Mills instantons. These are characterized by two integers which can be expressed as integrals quadratic in the Riemann tensor.

When quadratic terms are introduced into the GTG Lagrangian the theory differs markedly from similar extensions in GR. In GR one obtains fourth order equations for the metric [50], whereas in GTG one has a pair of lower order equations. One of these determines the connection, which in general will differ from that used in GR even in the absence of spin.

We construct the topological invariants for the GTG action integral. We show that the two invariants are the scalar and pseudoscalar parts of a single quantity, and our derivation

treats them in a unified way. The relationship with instanton solutions in Euclidean gravity is explored. As for instantons in Yang-Mills theory the rotation gauge field becomes pure gauge at infinity and the topological invariants are the corresponding winding numbers.

We construct irreducible fields from the Riemann tensor and use these to form ten independent quadratic terms from the Riemann tensor. In an action integral the two topological terms can be ignored, so only eight terms are needed. We construct the field equations for the parity non-violating Lagrangian terms. Units with  $\hbar = c = 8\pi G = 1$  are used throughout this chapter.

## 5.2 Topological invariants

We are interested in the behaviour of quadratic terms in the gravitational Lagrangian in GTG. We start by constructing the following quantity

$$\mathcal{Z} \equiv \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \underline{R}_{cd} \underline{R}_{ab} = \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \frac{1}{2} (\underline{R}_{cd} \underline{R}_{ab} + \underline{R}_{ab} \underline{R}_{cd}). \quad (5.1)$$

This only has scalar and pseudoscalar parts and can be written in terms of the non-covariant field strength as

$$\mathcal{Z} = \bar{h}^a \wedge \bar{h}^b \wedge \bar{h}^c \wedge \bar{h}^d \underline{R}_{cd} \underline{R}_{ab} = \det h \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \underline{R}_{cd} \underline{R}_{ab} \equiv \det h Z \quad (5.2)$$

where

$$Z \equiv \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \underline{R}_{cd} \underline{R}_{ab}. \quad (5.3)$$

We can now form an invariant integral that is independent of the  $\bar{h}^a$  field as

$$S \equiv \int |d^4x| \det h^{-1} \mathcal{Z} = \int |d^4x| Z. \quad (5.4)$$

From the definition of the Riemann tensor we find that

$$\begin{aligned} Z &= \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d (2\nabla_c \Omega_d + \Omega_c \Omega_d) (2\nabla_a \Omega_b + \Omega_a \Omega_b) \\ &= -4\gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \nabla (\nabla_c \Omega_a \Omega_b + \frac{1}{3} \Omega_a \Omega_b \Omega_c) \\ &= 2\gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \nabla (\underline{R}_{ac} \Omega_b + \frac{1}{3} \Omega_a \Omega_b \Omega_c). \end{aligned} \quad (5.5)$$

The main step in this derivation is the observation that the totally antisymmetrized product of 4 bivectors vanishes identically in 4-d. This proof that  $Z$  is a total divergence is considerably simpler than that given in [51], where gamma matrices were introduced in order to generate a similar ‘simple’ proof in the Riemann-Cartan formulation. Here we have also treated the scalar and pseudoscalar parts in a single term, which halves the work.

Since the integral reduces to a boundary term it should only contribute a global topological term to an action integral, and should not contribute to the local field equations. This is simple to check. There is no dependence on the  $\bar{h}^a$  field, so no contribution arises when this field is varied. When the  $\Omega_a$  field is varied one picks up terms proportional to

$$\gamma_a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d D_d \underline{R}_{cb} = \frac{1}{3} \gamma_a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d (D_d \underline{R}_{cb} + D_b \underline{R}_{dc} + D_c \underline{R}_{bd}) = 0, \quad (5.6)$$



which vanishes from the definition of the field strength and the Jacobi identity. Since the two topological terms do not contribute to the field equations, and can therefore be ignored in any classical action integral, it is useful to have expressions for them in terms of simpler combinations of the Riemann tensor and its contractions. For the scalar term we find that

$$\begin{aligned}
\langle \mathcal{Z} \rangle &= \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \mathcal{R}_{cd} \wedge \mathcal{R}_{ab} \\
&= (\gamma^a \wedge \gamma^b \wedge \gamma^c) \cdot [(\gamma^d \cdot \mathcal{R}_{cd}) \wedge \mathcal{R}_{ab} + \mathcal{R}_{cd} \wedge (\gamma^d \cdot \mathcal{R}_{ab})] \\
&= (\gamma^a \wedge \gamma^b) \cdot [-\mathcal{R} \mathcal{R}_{ab} + 2\mathcal{R}_c \wedge (\gamma^c \cdot \mathcal{R}_{ab}) + \mathcal{R}_{cd} (\gamma^c \wedge \gamma^d) \cdot \mathcal{R}_{ab}] \\
&= \mathcal{R}^2 + 2\gamma^a \cdot [\gamma^b \cdot \mathcal{R}_c \gamma^c \cdot \mathcal{R}_{ab} - \gamma^b \cdot (\gamma^c \cdot \mathcal{R}_{ab}) \mathcal{R}_c] + 2\mathcal{R}_{ba} \cdot \bar{\mathcal{R}}^{ab} \\
&= 2\mathcal{R}_{ba} \cdot \bar{\mathcal{R}}^{ab} - 4\mathcal{R}_a \cdot \bar{\mathcal{R}}^a + \mathcal{R}^2,
\end{aligned} \tag{5.7}$$

where the adjoint functions are defined by

$$(\gamma_a \wedge \gamma_b) \cdot \bar{\mathcal{R}}_{cd} \equiv (\gamma_c \wedge \gamma_d) \cdot \mathcal{R}_{ab} \quad \gamma_a \cdot \bar{\mathcal{R}}_b = \gamma_b \cdot \mathcal{R}_a. \tag{5.8}$$

For the pseudoscalar term (denoted  $\langle \mathcal{Z} \rangle_4$ ) we similarly obtain

$$\begin{aligned}
\langle \mathcal{Z} \rangle_4 &= \gamma^a \wedge \gamma^b \wedge \gamma^c \wedge \gamma^d \mathcal{R}_{cd} \cdot \mathcal{R}_{ab} \\
&= \gamma^a \wedge \gamma^b \wedge (\bar{\mathcal{R}}_{cd} (\gamma^c \wedge \gamma^d) \cdot \mathcal{R}_{ab}) \\
&= -I (\gamma^c \wedge \gamma^d) \cdot \mathcal{R}_{ab} (I \gamma^a \wedge \gamma^b) \cdot \bar{\mathcal{R}}_{cd} \\
&= 2I \mathcal{R}_{cd}^* \cdot \mathcal{R}^{cd}
\end{aligned} \tag{5.9}$$

where we have introduced the dual of the Riemann tensor defined by

$$\mathcal{R}_{ab}^* \equiv \frac{1}{2} I \gamma_a \wedge \gamma_b \wedge \gamma^d \wedge \gamma^c \mathcal{R}_{cd}. \tag{5.10}$$

We therefore have

$$S = \int |d^4x| \det h^{-1} (2\mathcal{R}_{ba} \cdot \bar{\mathcal{R}}^{ab} - 4\mathcal{R}_a \cdot \bar{\mathcal{R}}^a + \mathcal{R}^2 + 2I \mathcal{R}_{ab}^* \cdot \mathcal{R}^{ba}). \tag{5.11}$$

This generalizes the usual GR expressions for the topological invariants to the case where the Riemann tensor need not be symmetric, as in the case when there is torsion. Both of the scalar and pseudoscalar contributions can usually be ignored in the action integral. The standard GR expressions are recovered by setting  $\bar{\mathcal{R}}_{ab} = \mathcal{R}_{ab}$  and  $\bar{\mathcal{R}}_a = \mathcal{R}_a$ .

### 5.3 Relation to instantons

The derivation of topological terms in GTG has a Euclidean analogue, which gives rise to instanton winding numbers as found in Yang-Mills theory. For this section we assume that we are working in a Euclidean space. Most of the formulae go through unchanged, except that now the pseudoscalar squares to  $+1$ . For this section we therefore denote the pseudoscalar by  $E$ . The proof that the integral (5.4) is a total divergence is unaffected, and

so it can be converted to a surface integral. The Riemann is assumed to fall off sufficiently quickly that we can drop the  $R_{ac}$  term, so

$$S = -\frac{2}{3} \int |d^3x| n \wedge \gamma^a \wedge \gamma^b \wedge \gamma^c \Omega_a \Omega_b \Omega_c. \quad (5.12)$$

For the Riemann to tend to zero the  $\Omega_a$  field must tend to pure gauge,

$$\Omega_a = -2\nabla_a L \tilde{L}, \quad (5.13)$$

where  $L$  is a (Euclidean) rotor. The integral is invariant under continuous transformations of the rotor  $L$ , so we define the winding numbers

$$\chi + E\tau \equiv \frac{1}{6\pi^2} \int |d^3x| n \wedge \gamma^a \wedge \gamma^b \wedge \gamma^c \nabla_a L \tilde{L} \nabla_b L \tilde{L} \nabla_c L \tilde{L} = \frac{1}{32\pi^2} S. \quad (5.14)$$

The numbers  $\tau$  and  $\chi$  are instanton numbers for the solution, here given by the scalar and pseudoscalar parts of one equation. The common origin of the invariants is clear, as is the fact that one is a scalar and one a pseudoscalar. There are two integer invariants because the 4-d Euclidean rotor group is  $Spin(4)$  and the homotopy groups obey

$$\pi_3(Spin(4)) = \pi_3(SU(2) \times SU(2)) = \pi_3(SU(2)) \times \pi_3(SU(2)) = Z \times Z. \quad (5.15)$$

We can see this explicitly in GA, here the ESTA (Euclidean spacetime algebra), where

$$\gamma_\mu^2 = -1 \quad \sigma_k^2 = (E\sigma_k)^2 = -1 \quad E^2 = 1. \quad (5.16)$$

That the 4D rotor group is  $SU(2) \times SU(2)$  is closely related to the existence of an ideal structure for the bivectors. We can decompose a bivector  $F$  as

$$F = F_1 + F_2 = \frac{1}{2}(1 + E)\mathbf{A} + \frac{1}{2}(1 - E)\mathbf{B}. \quad (5.17)$$

This splits  $F$  into the sum of two commuting mutually annihilating bivectors  $F_1$  and  $F_2$ . A general rotor  $L$  can therefore be split into the product of two commuting rotors

$$L = e^F = e^{F_1} e^{F_2}. \quad (5.18)$$

Now

$$e^{\frac{1}{2}(1 \pm E)\mathbf{v}} = \frac{1}{2}(1 \pm E)e^{\mathbf{v}} + \frac{1}{2}(1 \mp E) \quad (5.19)$$

so

$$e^F = \frac{1}{2}(1 + E)e^{E\mathbf{A}} - \frac{1}{2}(1 - E)e^{E\mathbf{B}}. \quad (5.20)$$

Substituting for  $L$  in (5.14) we find that the integral splits into the sum of two ideals

$$\frac{1}{2}(\chi + E\tau) = \frac{1}{2}(1 + E)p_a + \frac{1}{2}(1 - E)p_b \quad (5.21)$$

where  $p_a$  and  $p_b$  are the winding numbers for the rotors  $e^{E\mathbf{A}}$  and  $e^{E\mathbf{B}}$ :

$$p_v = \frac{E}{12\pi^2} \int |d^3x| n \wedge \gamma^a \wedge \gamma^b \wedge \gamma^c \nabla_a e^{E\mathbf{v}} e^{-E\mathbf{v}} \nabla_b e^{E\mathbf{v}} e^{-E\mathbf{v}} \nabla_c e^{E\mathbf{v}} e^{-E\mathbf{v}} \quad (5.22)$$

The maps are clearly  $S^3 \rightarrow S^3$ , so we have the expected structure.

We can if we wish decompose the Euclidean rotation gauge function into the sum of two  $SU(2)$  gauge functions

$$\Omega_a = \frac{1}{2}(1 + E)\mathbf{A}_a + \frac{1}{2}(1 - E)\mathbf{B}_a \quad (5.23)$$

and come to the same conclusion.

The use of the ideal structure here is essentially the same as the complexification trick used to create commuting generators for the Lorentz group (see [52] and [53]). It is all a case of changing the structure from  $R^3 \times S^3$  (non-compact) to  $S^3 \times S^3$  (compact).

## 5.4 Topological terms for the $\bar{h}$ -field

We have found combinations of terms in the Riemann tensor that contribute a divergence to an action integral and therefore do not affect the local physics. Can we find similar terms involving the torsion tensor  $\mathcal{X}_a$ ? The answer is yes [54], some work shows that

$$\mathcal{X}_a \wedge \mathcal{X}^a - \mathcal{R}_{ab} \wedge \gamma^a \wedge \gamma^b = \det h \nabla \wedge [\bar{e}^a \wedge \underline{X}_a] \quad (5.24)$$

where  $\underline{X}_a = \bar{e}(\mathcal{X}_a) = D \wedge \bar{e}_a$  is the non-covariant torsion tensor. However, unlike the two quadratic Riemann terms, the quadratic torsion term is not scale invariant.

## 5.5 Quadratic Lagrangians

We now use the preceding results to construct a set of independent Lagrangian terms for GTG which are quadratic in the Riemann tensor  $\mathcal{R}_{ab}$ . The field equations that result will be very different to those obtained from quadratic GTG since in the gauge theory approach we consider  $\Omega_a$  as an independent field. We do not include quadratic terms in the torsion, largely because this work was motivated by the development of a scale invariant theory, and the quadratic torsion terms are not scale invariant.

To construct the independent terms for a quadratic Lagrangian we need to construct the irreducible parts of the Riemann tensor. To do this we write

$$\mathcal{R}_{ab} = \mathcal{W}_{ab} + \mathcal{P}_{ab} + \mathcal{Q}_{ab} \quad (5.25)$$

where

$$\gamma^a \mathcal{W}_{ab} = 0 \quad \gamma^a \mathcal{P}_{ab} = \gamma^a \wedge \mathcal{P}_{ab} \quad \gamma^a \cdot \mathcal{Q}_{ab} = \mathcal{R}_b. \quad (5.26)$$

In the language of Clifford analysis, this is a form of monogenic decomposition of  $\mathcal{R}_{ab}$  [52, 55]. To achieve this decomposition we start by defining [5]

$$\mathcal{Q}_{ab} = \mathcal{R}_{[a} \wedge \gamma_{b]} - \frac{1}{6} \gamma_a \wedge \gamma_b \mathcal{R}, \quad (5.27)$$

which satisfies  $\gamma^a \cdot \mathcal{Q}_{ab} = \mathcal{R}_b$ . We next take the protraction of (5.25) with  $\gamma^a$  to obtain

$$\gamma^a \wedge \mathcal{R}_{ab} - \frac{1}{2} \gamma^a \wedge \mathcal{R}_a \wedge \gamma_b = \gamma^a \wedge \mathcal{P}_{ab}. \quad (5.28)$$

We now define the vector valued function

$$\mathcal{V}_b \equiv -I\gamma^a \wedge \mathcal{R}_{ab} = \gamma^a \cdot (I\mathcal{R}_{ab}). \quad (5.29)$$

The symmetric part of  $\mathcal{V}_b$  is

$$\begin{aligned} \mathcal{V}_b^+ &= \frac{1}{2}(\mathcal{V}_b + \gamma^a \mathcal{V}_a \cdot \gamma_b) \\ &= -I\frac{1}{2}(\gamma^a \wedge \mathcal{R}_{ab} + \gamma^a \gamma_b \wedge \gamma^c \wedge \mathcal{R}_{ca}) \\ &= -I(\gamma^a \wedge \mathcal{R}_{ab} - \frac{1}{2}\gamma^a \wedge \mathcal{R}_a \wedge \gamma_b) \end{aligned} \quad (5.30)$$

so we have

$$\gamma^a \wedge \mathcal{P}_{ab} = I\mathcal{V}_b^+. \quad (5.31)$$

It follows that

$$\mathcal{P}_{ab} = -I\mathcal{V}_{[a}^+ \wedge \gamma_{b]} + \frac{1}{6}I\gamma_a \wedge \gamma_b \mathcal{V} \quad (5.32)$$

where

$$\mathcal{V} = \gamma^a \cdot \mathcal{V}_a. \quad (5.33)$$

This construction of  $\mathcal{P}_{ab}$  ensures that the tensor has zero contraction, as required.

Splitting the Ricci tensor into symmetric and antisymmetric parts we can finally write the Riemann tensor as

$$\mathcal{R}_{ab} = \mathcal{W}_{ab} + \mathcal{R}_{[a}^+ \wedge \gamma_{b]} - \frac{1}{6}\gamma_a \wedge \gamma_b \mathcal{R} + \mathcal{R}_{[a}^- \wedge \gamma_{b]} - I\mathcal{V}_{[a}^+ \wedge \gamma_{b]} + \frac{1}{6}I\gamma_a \wedge \gamma_b \mathcal{V} \quad (5.34)$$

where  $+$  and  $-$  superscripts denote the symmetric and antisymmetric parts of a tensor respectively. This decomposition splits the Riemann tensor into a Weyl term ( $\mathcal{W}_{ab}$ ) with 10 degrees of freedom, two symmetric tensors ( $\mathcal{R}_a^+$  and  $\mathcal{V}_a^+$ ) with 10 degrees of freedom each, and an anti-symmetric tensor ( $\mathcal{R}_a^-$ ) with 6 degrees of freedom. These account for all 36 degrees of freedom in  $\mathcal{R}_{ab}$ . The first three terms in the decomposition are the usual ones for a symmetric Riemann tensor and would be present in GR. The remaining terms come from the antisymmetric parts of  $\mathcal{R}_{ab}$  and only arise in the presence of spin or quadratic terms in the Lagrangian. It is now a simple task to construct traceless tensors from  $\mathcal{V}_a^+$  and  $\mathcal{R}_a^+$  to complete the decomposition into irreducible parts.

We can write the antisymmetric part of  $\mathcal{R}_a$  as

$$\mathcal{R}_a^- = \gamma_a \cdot \mathcal{A} \quad (5.35)$$

where  $\mathcal{A} = \frac{1}{2}\gamma^a \wedge \mathcal{R}_a$  is a bivector. Using this definition we can write down 10 independent scalar terms which are quadratic in the Riemann tensor:

$$\begin{aligned} &\{ \mathcal{W}^{ab} \cdot \mathcal{W}_{ab}, \quad \mathcal{W}^{ab} \cdot (I\mathcal{W}_{ab}), \quad \mathcal{R}^{+a} \cdot \mathcal{R}_a^+, \quad \mathcal{R}^2, \\ &\mathcal{A} \cdot \mathcal{A}, \quad \mathcal{A} \cdot (I\mathcal{A}), \quad \mathcal{V}^{+a} \cdot \mathcal{V}_a^+, \quad \mathcal{V}^{+a} \cdot \mathcal{R}_a^+, \quad \mathcal{V}^2, \quad \mathcal{R}\mathcal{V} \} \end{aligned} \quad (5.36)$$

Six of these are invariant under parity and four are parity violating. The two topological invariants can be used to remove two terms, so there are only eight possible independent quadratic terms for the gravitational Lagrangian. The classical field equations arising from

an equivalent set of terms, in addition to the possible torsion squared terms, are calculated in [56]. We shall briefly give the non-parity violating field equations here.

For calculational purposes it is easier to use the six parity invariant terms

$$\{ \mathcal{R}^{ab} \cdot \mathcal{R}_{ba}, \quad \mathcal{R}^a \cdot \mathcal{R}_a, \quad \bar{\mathcal{R}}^a \cdot \mathcal{R}_a, \quad \mathcal{R}^2, \quad \mathcal{V}^a \cdot \mathcal{V}_a, \quad \mathcal{V}^2 \} \quad (5.37)$$

and the four parity violating terms

$$\{ \mathcal{R}^{ab} \cdot (I\mathcal{R}_{ba}), \quad \mathcal{R}^a \cdot \mathcal{V}_a, \quad \bar{\mathcal{R}}^a \cdot \mathcal{V}_a, \quad \mathcal{R}\mathcal{V} \} \quad (5.38)$$

which are linear combinations of the irreducible components. For example the Weyl squared term is given by

$$\mathcal{W}_{ab} \cdot \mathcal{W}^{ba} = \mathcal{R}_{ab} \cdot \mathcal{R}^{ba} - \frac{1}{2} \mathcal{R}_a \cdot \mathcal{R}^a - \frac{1}{2} \bar{\mathcal{R}}_a \cdot \mathcal{R}^a + \mathcal{V}_a \cdot \mathcal{V}^a + \frac{1}{6} (\mathcal{R}^2 + \mathcal{V}^2). \quad (5.39)$$

The topological invariants can be used to remove one term from each set. If we consider just the parity invariant terms and use the topological invariant to remove  $\mathcal{R}^a \cdot \mathcal{R}_a$  we can calculate the field equations from

$$\mathcal{L}_{R^2} = \frac{1}{4} \epsilon_1 \mathcal{R}^2 + \frac{1}{2} \epsilon_2 \mathcal{R}^a \cdot \mathcal{R}_a + \frac{1}{4} \epsilon_3 \mathcal{R}^{ab} \cdot \mathcal{R}_{ba} + \epsilon_4 \frac{1}{4} \mathcal{V}^2 + \epsilon_5 \frac{1}{2} \mathcal{V}^a \cdot \mathcal{V}_a \quad (5.40)$$

The field equations for the  $h^a$  give a modified Einstein tensor of the form

$$\mathcal{G}'_a = \mathcal{G}_a + \epsilon_1 \mathcal{G}_{1a} + \epsilon_2 \mathcal{G}_{2a} + \epsilon_3 \mathcal{G}_{3a} + \epsilon_4 \mathcal{G}_{4a} + \epsilon_5 \mathcal{G}_{5a} \quad (5.41)$$

where

$$\mathcal{G}_{1a} = \mathcal{R}(\mathcal{R}_a - \frac{1}{4} \gamma_a \mathcal{R}) \quad (5.42)$$

$$\mathcal{G}_{2a} = \gamma_b \mathcal{R}^b \cdot \mathcal{R}_a + \mathcal{R}_{ab} \cdot \mathcal{R}^b - \frac{1}{2} \gamma_a \mathcal{R}^b \cdot \mathcal{R}_b \quad (5.43)$$

$$\mathcal{G}_{3a} = \gamma_b \mathcal{R}^{bc} \cdot \mathcal{R}_{ca} - \frac{1}{4} \gamma_a \mathcal{R}^{bc} \cdot \mathcal{R}_{cb} \quad (5.44)$$

$$\mathcal{G}_{4a} = \mathcal{V}(\mathcal{V}_a - \frac{1}{4} \gamma_a \mathcal{V}) \quad (5.45)$$

$$\mathcal{G}_{5a} = \gamma_b \mathcal{V}^b \cdot \mathcal{V}_a + (I\mathcal{R}_{ab}) \cdot \mathcal{V}^b - \frac{1}{2} \gamma_a \mathcal{V}^b \cdot \mathcal{V}_b. \quad (5.46)$$

These tensors all have zero contraction, as expected from scale invariance.

The field equations for  $\Omega_a$  give the generalized structure equation of the form

$$\mathcal{N}'_a = \mathcal{S}_a \quad (5.47)$$

where  $\mathcal{S}_a$  is the covariant matter spin tensor. The contributions to  $\mathcal{N}'_a$  from the five terms in the action integral are then given by

$$\mathcal{N}'_{1a} = \mathcal{R} \gamma^b \cdot (\gamma_a \wedge \mathcal{X}_b) + \gamma_a \wedge \mathcal{D}\mathcal{R} \quad (5.48)$$

$$\mathcal{N}'_{2a} = ((\gamma^c \wedge \gamma^b) \cdot (\gamma_a \wedge \mathcal{X}_c)) \wedge \mathcal{R}_b + \gamma_a \wedge (\mathcal{D}_b \mathcal{R}^b) - \mathcal{D} \wedge \mathcal{R}_a \quad (5.49)$$

$$\mathcal{N}'_{3a} = \mathcal{D}^b \mathcal{R}_{ab} - \mathcal{X}^b \mathcal{R}_{ab} + \frac{1}{2} \mathcal{R}_{bc} \mathcal{X}^{bc} a \quad (5.50)$$

$$I\mathcal{N}'_{4a} = \mathcal{V} \gamma^b \cdot (\gamma_a \wedge \mathcal{X}_b) + \gamma_a \wedge \mathcal{D}\mathcal{V} \quad (5.51)$$

$$I\mathcal{N}'_{5a} = ((\gamma^c \wedge \gamma^b) \cdot (\gamma_a \wedge \mathcal{X}_c)) \wedge \mathcal{V}_b + \gamma_a \wedge (\mathcal{D}_b \mathcal{V}^b) - \mathcal{D} \wedge \mathcal{V}_a \quad (5.52)$$

where  $\mathcal{X}^b = \gamma^b \cdot \mathcal{X} = \gamma^b \cdot (\gamma^a \cdot \mathcal{X}_a)$ . Each of these terms is locally scale covariant, as they must be coming from a scale invariant Lagrangian. As for the Dirac Lagrangian the local scale covariance is maintained by the torsion tensor —  $\mathcal{X}$  enters in just the right way to form scale covariant combinations with the covariant derivatives.

Either these terms have to be significant only at high energy, or they have to be consistent with GR. For a vacuum solution of GR we have  $\mathcal{R}_{ab} = \bar{\mathcal{R}}_{ab}$ ,  $\gamma^a \mathcal{R}_{ab} = 0$  and  $\mathcal{R}_a = 0$ . It follows that in this case all the quadratic terms in the action are zero or a total divergence (The  $\mathcal{R}_{ab} \cdot \mathcal{R}^{ba}$  term is a divergence from (5.11)). Vacuum solution to GR are therefore also solutions to a theory with quadratic curvature terms. The quadratic terms will also be small when the curvature is small.

With a Ricci plus Weyl squared Lagrangian the modified structure equation is automatically satisfied in the absence of torsion and spin. Isotropic cosmological solutions of GR (which have zero Weyl tensor) will therefore also be solutions of this theory. Actions involving terms quadratic in the torsion will also trivially be solved by any torsion-free solutions to GR, however as previously noted these are not scale invariant. For a variety of more general results on the correspondence with GR see [56].

## 5.6 Conclusions

We have shown that in gauge theory gravity topological terms are simply dealt with and reduce to boundary integrals which do not alter the classical field equations. In the gauge theory approach these topological terms are the winding numbers for instanton solutions in Euclidean gravity. We constructed ten possible terms for a quadratic Lagrangian, which the topological invariants then restrict to eight independent terms, and derived the field equations. These terms may make important contributions to the high-energy gravity theory, with or without the motivation of being valid contributions to a scale invariant action.

## Chapter 6

# Perturbations in Cosmology

In this chapter we briefly review our current understanding of the history of the universe and then move on to consider covariant methods for analysing cosmological perturbations within that framework.

We start by briefly describing the relevant features of the big bang model, covering the early inflationary period through to the universe we observe today. The general model is well established and is consistent with current observations (for a general introduction see [57, 58]; for an alternative viewpoint see [46]). On large scales the universe appears to be approximately homogeneous and isotropic. Solutions to General Relativity for an exactly homogeneous and isotropic matter distribution form an exact family of theoretical models, the Friedmann-Robertson-Walker (FRW) universes. In reality the universe is not exactly FRW and so we move on to discuss perturbations about an exact FRW universe. These perturbations leave their imprint on the Cosmic Microwave Background (CMB) radiation and ultimately evolve to form the structure that we observe today.

In the spirit of this thesis we employ covariant methods to study the perturbations, so ensuring that the quantities we discuss are directly observable and there are no ambiguities inherited from the gauge freedom in the underlying gravitational theory. The main drawback to the covariant approach is that it hides the correspondence with the gauge fields and so complicates any quantum considerations. We shall not attempt to surmount this difficulty here.

The presentation draws heavily on references [6–10] and other discussions can be found in [59–61]. Non-covariant techniques are well known and are described in many places, e.g. [12, 14]. Ultimately the linearized equations that we arrive at are equivalent to the equations derived by non-covariant techniques.

Much of this chapter is a review of previous work. The main innovations are in the derivation of the scalar-field inflation equations, the energy integrated multipole equations, and the equations governing the evolution of massive neutrino perturbations. These are innovations of presentation only, not in content (which is well known). The energy-integrated equations have previously been used to study small velocity dispersion of cold dark matter perturbations [62]. Here we derive the general equations which can be applied to all species of matter present and discuss how they can be used to study massive neutrino perturba-

tions. The material in this chapter forms the theoretical basis for much of the numerical implementation we shall discuss in the next chapter.

## 6.1 Introduction: Evolution of the Universe

Evidence from the redshift of objects observed by astronomers, and a body of other evidence, lead us to believe that the universe is currently expanding. Assuming that the universe has been expanding for all of its history we conclude that it must have started very small. Observations on the largest scales also lead us to conclude that the universe is approximately homogeneous and isotropic. Solutions to General Relativity for a homogeneous and isotropic medium are consistent with the apparent expansion and define the exact FRW universe models.

If we define the scale factor of the universe  $S$ , equal to one today, volumes in the past scale with  $S^3$ . Matter densities are therefore proportional to  $S^{-3}$ . Radiation is redshifted by the expansion and has energy proportional to  $S^{-1}$ , so the radiation energy density varies as  $S^{-4}$ . The radiation energy density today is very low, as witnessed by the approximately uniform microwave background temperature of 2.7K, so the evolution is matter dominated. However if we go back far enough the radiation density will come to dominate, and we enter the radiation dominated era.

At the moment the universe is only slightly ionized and is therefore transparent on large scales. However at approximately the same scale factor at which the universe becomes radiation dominated the photons have enough energy to ionize the hydrogen and helium that make up the bulk of the baryonic mass of the universe. This is called recombination (a misnomer). Before recombination there was a high free electron density and the universe was opaque. The CMB we observe in the sky today is therefore a snapshot of the time of recombination when the universe first became transparent. The surface we see in the CMB is called the last scattering surface, a spherical shell of the universe in its early stages of evolution.

As we go back in time before recombination the age of the universe becomes so short that the distance light could have travelled becomes very short. At some point in the past all perturbations at astronomically relevant scales today would have been out of causal contact. This provides the ‘Horizon Problem’— we observe correlations on scales that could not have been in causal contact at the beginning of an FRW universe containing only normal matter. The currently favoured solution to this is an inflationary scenario in which there was an early period of exponential expansion. Scales we observe today would then once again be in causal contact at the early stage of inflation and the horizon problem goes away. Approximately exponential expansion can be obtained by postulating a scalar ‘inflaton’ field that dominated the very early universe.

Inflation also provides a framework in which perturbations can be generated by quantum (or thermal) fluctuations during the exponential expansion. These get rapidly pushed outside the Hubble radius (the radius of causal contact) by the expansion, and only re-enter the horizon again at some point during the radiation or matter dominated eras. Once perturbations are outside the horizon they cannot be greatly affected by local physics, and



we expect the perturbations re-entering the horizon to have approximately the same form as when they left it. This allows us to make concrete predictions for the early radiation dominated era ignoring all the unknown complications of reheating (the generation of the matter in the universe today at the end of inflation).

There are three families of FRW universes corresponding to their spatial curvatures (which are determined by the total energy density and expansion rate). In a ‘flat’ universe the 3-space is Euclidean so light rays travel in straight lines. In ‘open’ and ‘closed’ universes light rays diverge or converge respectively. Whether the curvature will be important depends on whether there has been sufficient time for deviations from flatness to be observed, corresponding to the relative sizes of the Hubble radius and curvature radius. Evidence suggests that the universe is nearly flat today, so in the early universe the curvature would have been insignificant. However the angular size at which we see scales on the last scattering surface will depend on the curvature. For example in a closed model the light rays will have converged, so a given angle on the sky today corresponds to a smaller size on the last scattering surface than in open or flat models. The curvature must therefore be taken into account when discussing the late evolution. The observed angular sizes of scales in the CMB provide a good way to measure the curvature.

In the radiation and matter dominated eras perturbations on scales larger than the Hubble radius can be viewed as little mini-universes with a larger or smaller energy density than the background. An over-dense region therefore evolves as a more closed universe and an under-dense region as a more open universe. Closed universes expand less fast than flat ones, so as they evolve an over-dense region will become even more over-dense relative to the background, and so the over-density increases. Similarly under-dense regions grow faster than the background becoming more under-dense. Thus perturbations on super-Hubble scales grow as the universe evolves. When the perturbation enters the horizon it can undergo gravitational collapse. The perturbation collapses until the pressure dominates, and then a series of oscillations result during which the perturbation amplitude can no longer grow due to the pressure support. The oscillations are damped by photons diffusing from hot to cold regions, and small scale perturbations can be virtually wiped out over time. What we see on the last scattering surface at a particular angular size will depend on the stage of evolution of the corresponding perturbation scale at the time of recombination.

Observations of the CMB show that it is not exactly uniform at 2.7K but that there are anisotropies at the  $10^{-5}$  level. These resulted from perturbations in the baryonic matter and radiation on the last scattering surface. The smallness of the CMB anisotropy is very helpful as it implies that a linearized treatment, which greatly simplifies the analysis, will be highly accurate at least until last scattering.

Perturbations in the total matter of order  $10^{-5}$  are not sufficient to explain all the small scale structure we observe today. This can be solved by the introduction of a cold dark matter (CDM) component. CDM is any matter with negligible velocity dispersion that does not couple to the baryons or photons and therefore evolves purely under the influence of gravity. While the baryonic perturbations are prevented from growing due to pressure support (and decay due to damping) the CDM perturbations can continue to grow. At recombination we see an anisotropy in the CMB due to a small baryonic/photon perturbation. However after recombination atoms have formed, the baryons are no longer tightly coupled to the

photons, and so the photon pressure is unimportant. The baryons can then fall into the potential wells of the CDM perturbations. Thus the ultimate total matter perturbation after last scattering can be considerably greater than the anisotropy we observe in the CMB, allowing the  $10^{-5}$  figure to be consistent with structures evolved from larger total matter perturbations. Dark matter can also provide an explanation for other astronomical observations such as the shape of galaxy rotation curves, though recent evidence is beginning to suggest that CDM may not be sufficient to explain all the observations. The situation may be helped by the introduction of hot dark matter (massive neutrinos), warm dark matter or interacting dark matter components [63–65], or even by using a different gravitational theory [66].

General Relativity admits a cosmological constant,  $\Lambda$ , corresponding to a uniform energy density with isotropic negative pressure. As  $\Lambda$  is a constant it will be dominated by the radiation and matter density in the early universe if it is to have values consistent with observations today. Current evidence suggests that the expansion of the universe is accelerating and therefore that we have just entered the  $\Lambda$ -dominated era. In the  $\Lambda$ -dominated era the acceleration due to the negative pressure overcomes the deceleration we expect from the gravitational attraction of the matter and radiation. One can also devise theories in which the acceleration is explained by a ‘quintessence’, where the negative pressure comes from a scalar field background that is evolving with time. We do not discuss quintessence models here, but do (trivially) allow for a cosmological constant via a homogeneous stress-energy component with  $\rho = -p$ .

During inflation the universe is also accelerating. However since we know inflation has to come to an end it is most readily explained by the presence of a slowly evolving scalar field rather than a strictly constant  $\Lambda$  term.

The spatial curvature is commonly measured by  $\Omega_{\text{tot}}$  which is the ratio of the total energy density today to that required for a flat universe. If the universe never enters a phase in which it is  $\Lambda$ -dominated the curvature also determines the fate of the universe—closed models  $\Omega_{\text{tot}} > 1$  will recollapse to a ‘big crunch’ and open models ( $\Omega_{\text{tot}} < 1$ ) will go on expanding for ever. Once the universe enters a  $\Lambda$ -dominated phase it will accelerate for ever regardless of the spatial curvature.

We now discuss in detail the propagation of classical linear perturbations and apply the equations to perturbations during inflation. We demonstrate exactly which quantity we expect to be preserved for super-Hubble modes over the reheating epoch. We then consider the era after reheating from the early radiation dominated era, when all relevant scales are still outside the horizon, until the present day. The main goal of this is to be able to calculate the power spectrum for CMB anisotropies observed on the sky today in terms of the super-Hubble perturbations present in the early radiation dominated era. Since inflation is rather speculative we allow for an arbitrary initial power spectrum rather than fixing it to the predictions from a particular inflationary model. With the numerical implementation in the next chapter we should ultimately be able to use forthcoming observations to constrain the initial power spectrum and hence the inflationary theory.

## 6.2 Covariant perturbations: The 3+1 decomposition

We can perform a space-time split of physical quantities with respect to a four velocity  $u$ , where it is convenient to choose  $u^2 = 1$ . The component quantities are then in principle observable by observers moving with velocity  $u$ . In an exact FRW universe we can unambiguously choose  $u$  to be the velocity of fundamental observers (observers comoving with the matter). However in a perturbed universe there is freedom to choose  $u$  in many ways such that it reduces to the velocity of a fundamental observer in the FRW limit. It is often convenient to choose  $u$  such that particular quantities vanish, maybe the total energy flux (the energy frame), the shear (the Newtonian gauge) or the CDM velocity (the CDM frame). For the moment we shall leave  $u$  unspecified.

The divergence of the velocity, the expansion scalar  $\theta \equiv \mathcal{D} \cdot u$ , will be a constant in the 3-volume orthogonal to  $u$  in an isotropic universe. If we consider the evolution of a 3-volume along the lines of  $u$  we can use the divergence theorem and  $u^2 = 1$  to write

$$\int dV_4 \mathcal{D} \cdot u = \int dt V_3 \theta = \Delta V_3 \quad (6.1)$$

and hence

$$\theta = \frac{1}{V} \frac{dV_3}{dt} = 3 \frac{\dot{S}}{S}. \quad (6.2)$$

Here  $S$  is the scale factor so that  $V_3 \propto S^3$ , an over-dot denotes the time derivative  $u \cdot \mathcal{D}$ , and we can choose to normalise so that  $S = 1$  today. The Hubble parameter is often used to discuss the expansion rate and is defined as  $\dot{S}/S = \theta/3$ .

It is useful to define a projection operator to project quantities orthogonal to  $u$

$$\underline{H}(a) = a - u \cdot a u. \quad (6.3)$$

We also define the scalar

$$h_{ab} \equiv \gamma_a \cdot \underline{H}_b = h_{ba} \quad (6.4)$$

and then  $\underline{H}_a u^a = 0$  and  $h_b^a \underline{H}_a = \underline{H}_b$  as required for a projection orthogonal to  $u$ . (In this chapter we concentrate on covariant methods and do not use the  $h_{ab} = \gamma_a \cdot h_b$  function so no confusion should arise).

We can write a general covariant tensor  $\mathcal{V}_{ab} = \gamma_a \cdot \underline{\mathcal{V}}_b$  in terms of irreducible components

$$\mathcal{V}_{ab} = \mathcal{V}_{\langle ab \rangle} + \mathcal{V}_{[ab]} + 2u^c \mathcal{V}_{c(a} u_{b)} + u^c u^d \mathcal{V}_{cd} u_a u_b - \frac{1}{3} h^{cd} \mathcal{V}_{cd} h_{ab}. \quad (6.5)$$

The angled brackets denote the projected symmetric trace free (PSTF) part, square brackets antisymmetrization and round brackets symmetrization. The antisymmetric part corresponds to a bivector that can be further split into one even and one odd parity spin-1 fields.

In using the 3+1 decomposition we often equate irreducible parts on both sides on an equation. The product of two irreducible tensors is not in general irreducible, the PSTF part of the product of a PSTF tensor  $F_{a_1 \dots a_l}$  and a projected vector  $J_a$  is given by

$$J_{\langle a} F_{a_1 \dots a_l \rangle} = J_{(a} F_{a_1 \dots a_l)} - \frac{l}{2l+1} J^b F_{b(a_1 \dots a_{l-1}} h_{a_l a)} \quad (6.6)$$

which can be verified by showing that the contraction on any pair of indices is zero. This is useful for analysing the PSTF parts of an equation and for building up higher rank PSTF tensors from lower rank ones.

The stress-energy tensor appears in Einstein's equation and is therefore of fundamental importance. We assume that it is symmetric (no spin) so that it can be decomposed as

$$\mathcal{T}_{ab} = \pi_{ab} + 2q_{(a}u_{b)} + \rho u_a u_b - p h_{ab} \quad (6.7)$$

where  $\pi_{ab}$  is PSTF and represents the anisotropic stress,  $q$  is the vector heat flux,  $p$  is the isotropic pressure and  $\rho$  is the energy density.

We also need to take derivatives in the 3-space orthogonal to  $u$ . For a tensor  $\mathcal{V}_b$  we define the projected derivative

$$\hat{\mathcal{D}}_a \mathcal{V}_b = h_a^c h_b^d H(\mathcal{D}_c \mathcal{V}_d) \quad (6.8)$$

or equivalently

$$\hat{\mathcal{D}}_a \mathcal{V}_{bc} = h_a^d h_b^e h_c^f \mathcal{D}_d \mathcal{V}_{ef} \quad (6.9)$$

The other important quantity is the covariant derivative of  $u$ , which is decomposed as

$$\mathcal{D}_a u_b = \sigma_{ab} + \varpi_{ab} + \frac{1}{3}\theta h_{ab} + u_a A_b \quad (6.10)$$

where  $\sigma_{ab} = \sigma_{(ab)}$  is the PSTF shear tensor,  $\theta = \mathcal{D} \cdot u$  is the expansion scalar,  $A = u \cdot \mathcal{D}u$  is the acceleration and  $\varpi_{ab} = \hat{\mathcal{D}}_{[a} u_{b]}$  can be written in terms of the vorticity bivector

$$\varpi_{ab} = \frac{1}{2}(\gamma_b \wedge \gamma_a) \cdot \varpi. \quad (6.11)$$

All the components of the derivative are orthogonal to  $u$  because  $u^2 = 1$ . The antisymmetric part of  $u_a A_b$  cancels with the vorticity to leave the antisymmetric part of the derivative

$$\mathcal{D} \wedge u = \varpi - A \wedge u. \quad (6.12)$$

From the definition the vorticity obeys the constraint equation

$$\hat{\mathcal{D}} \wedge \varpi + A \wedge \varpi = 0. \quad (6.13)$$

In an exact FRW universe all projected non-isotropic quantities vanish otherwise they could be used to define a direction in contradiction with the assumed isotropy. The only non-zero parts of the stress-energy tensor are therefore the energy density and isotropic pressure. Likewise the expansion scalar completely determines the derivative of  $u$ . All the other terms are therefore covariant first order quantities which vanish in an exact FRW model. Other quantities that do not vanish necessarily in an exact FRW universe are the spatial curvature and the scale factor  $S$  (in a perturbed universe defined by  $3\dot{S}/S = \theta$ ). However we can take the spatial gradient of any zero order scalar to define projected vectors that will be first order about an FRW model. We therefore quantify the perturbations about the exact zero-order FRW values with the variables

$$\mathcal{Z}_a \equiv S \hat{\mathcal{D}}_a \theta \quad h_a \equiv \hat{\mathcal{D}}_a S \quad \mathcal{X}_a \equiv S \hat{\mathcal{D}}_a \rho \quad \mathcal{X}_a^p \equiv S \hat{\mathcal{D}}_a p \quad (6.14)$$

and

$$\eta_a \equiv \frac{1}{2} S \hat{\mathcal{D}}_a {}^{(3)}\mathcal{R}. \quad (6.15)$$

To relate the time derivatives of these derived first order quantities to the time derivatives of the zero order quantities we can use the result for a scalar  $\phi$  that

$$(\hat{\mathcal{D}}_a \phi)^\cdot = \hat{\mathcal{D}}_a \dot{\phi} - \hat{\mathcal{D}}_a u_b \hat{\mathcal{D}}^b \phi - A_a \dot{\phi} - u_a A^\cdot \hat{\mathcal{D}} \phi. \quad (6.16)$$

To first order the  $u_a$  component vanishes and the time derivative is also projected.

This completes the set of variables useful for covariantly describing the evolution of perturbations about an FRW model. All the first order quantities are useful but they are not all independent, for example

$$\dot{h}_a = \frac{1}{3} (\mathcal{Z}_a - S \theta A_a). \quad (6.17)$$

Using the decomposition of the stress-energy tensor the equation for conservation of stress-energy  $\mathcal{D} \cdot \underline{\mathcal{T}}_a = 0$  implies the energy conservation equation

$$\dot{\rho} + (\rho + p)\theta + \hat{\mathcal{D}} \cdot q = 0 \quad (6.18)$$

and the momentum conservation equation

$$\dot{q}_a + \frac{4}{3} \theta q_a + (\rho + p) A_a - \hat{\mathcal{D}}_a p + \hat{\mathcal{D}} \cdot \underline{\pi}_a = 0. \quad (6.19)$$

The first order propagation equation for the density perturbation  $\mathcal{X}_a$  follows from taking the spatial gradient of (6.18) and commuting derivatives using (6.16) giving

$$\dot{\mathcal{X}}_a + 3\dot{h}_a(\rho + p) + (\mathcal{X}_a + \mathcal{X}_a^p)\theta + S \hat{\mathcal{D}}_a \hat{\mathcal{D}} \cdot q = 0. \quad (6.20)$$

The remaining physics governing the evolution of perturbations is given by Einstein's equation and the Bianchi identity, which we consider next.

### 6.3 General Relativity

To make progress we need to relate the variables in the stress-energy tensor to those in the derivative of  $u$  using Einstein's equation

$$\mathcal{R}_a - \frac{1}{2} \gamma_a \mathcal{R} = \kappa \underline{\mathcal{T}}_a. \quad (6.21)$$

We decompose the Riemann tensor as

$$\mathcal{R}_{ab} = \mathcal{W}_{ab} + \mathcal{R}_{[a} \wedge \gamma_{b]} - \frac{1}{6} \gamma_a \wedge \gamma_b \mathcal{R} \quad (6.22)$$

where  $\mathcal{W}_{ab}$  is the Weyl tensor which is symmetric and traceless. Einstein's equation then allows us to replace all the terms except the Weyl tensor with terms from the stress-energy tensor:

$$\mathcal{R}_{ab} = \mathcal{W}_{ab} + \kappa \underline{\mathcal{T}}_{[a} \wedge \gamma_{b]} - \kappa \frac{1}{3} \gamma_a \wedge \gamma_b \underline{\mathcal{T}} \quad (6.23)$$

where the trace of the stress-energy tensor is  $\mathcal{T} = \rho - 3p$ . To relate this to the terms in the decomposition of the derivative of  $u$  we use the Ricci equation

$$[\mathcal{D}_a, \mathcal{D}_b]u = \mathcal{R}_{ab} \cdot u. \quad (6.24)$$

To simplify matters we now only keep terms which are first order about an FRW universe. Exactly the same method can be used to obtain the exact equations in a straightforward way [6], but here we only require the linearized equations since observations tell us that the CMB anisotropies are small. Using the Ricci equation contracted with  $u^b$  we have

$$u^b u \cdot \mathcal{R}_{ab} = \frac{1}{3}(\dot{\theta} + \frac{1}{3}\theta^2)\underline{H}_a + \frac{1}{3}\theta\gamma_a \cdot \varpi + \frac{1}{2}\gamma_a \cdot \dot{\varpi} + \dot{\sigma}_a + \frac{2}{3}\theta\sigma_a - \hat{\mathcal{D}}_a A. \quad (6.25)$$

Note that to first order time derivatives of projected first order quantities (like  $\sigma_a$ ) are also projected since

$$u \cdot \dot{\sigma}_a = -\dot{u} \cdot \sigma_a = -A \cdot \sigma_a = 0 \quad (6.26)$$

to first order.

The quantity  $u^b \mathcal{W}_{ab} u$  can be split into vector and trivector parts

$$u^b \mathcal{W}_{ab} u \equiv \mathcal{E}_a + I\mathcal{B}_a. \quad (6.27)$$

defining the “electric” and “magnetic” parts of the Weyl tensor. The PSTF nature of these follows from the symmetry properties of the Weyl tensor and they are first order since they are projected vectors. Using the decomposition of the Riemann tensor we therefore have

$$u^b u \cdot \mathcal{R}_{ab} = -\mathcal{E}_a - \frac{1}{2}\kappa\pi_a - \frac{1}{6}\kappa H_a(\rho + 3p). \quad (6.28)$$

Equating the two expressions for  $u^b \mathcal{R}_{ab} \cdot u$  we have

$$\begin{aligned} \frac{1}{3}(\dot{\theta} + \frac{1}{3}\theta^2)\underline{H}_a + \frac{1}{3}\theta\gamma_a \cdot \varpi + \frac{1}{2}\gamma_a \cdot \dot{\varpi} + \dot{\sigma}_a + \frac{2}{3}\theta\sigma_a - \hat{\mathcal{D}}_a A \\ = -\mathcal{E}_a - \frac{1}{2}\kappa\pi_a - \frac{1}{6}\kappa H_a(\rho + 3p). \end{aligned} \quad (6.29)$$

We now equate the irreducible parts on both sides of the equation. Taking the trace gives the linearized Raychaudhuri equation

$$\dot{\theta} + \frac{1}{3}\theta^2 - \hat{\mathcal{D}} \cdot A = -\frac{1}{2}\kappa(\rho + 3p). \quad (6.30)$$

The antisymmetric part gives a propagation equation for the vorticity

$$\dot{\varpi} + \frac{2}{3}\theta\varpi - \hat{\mathcal{D}} \wedge A = 0 \quad (6.31)$$

and the PSTF part gives a propagation equation for the shear

$$\dot{\sigma}_{ab} + \frac{2}{3}\theta\sigma_{ab} - \hat{\mathcal{D}}_{\langle a} A_{b\rangle} + \mathcal{E}_{ab} + \frac{1}{2}\kappa\pi_{ab} = 0. \quad (6.32)$$

Taking the trace of the Ricci equation to give  $-u \cdot R_a u^a$  and using the Einstein equation leads to the constraint equation

$$\kappa q_a = \hat{\mathcal{D}}^b \varpi_{ab} + \hat{\mathcal{D}} \cdot \sigma_a - \frac{2}{3}\hat{\mathcal{D}}_a \theta. \quad (6.33)$$

Decomposing the dual of the Riemann tensor and using Einstein's equation gives

$$u^b \mathcal{R}_{ab}^* \cdot u = -\mathcal{B}_a + \frac{1}{2} \kappa I \gamma_a \wedge u \wedge q. \quad (6.34)$$

Applying the Ricci equation to the left hand side and taking the PSTF part yields the constraint equation

$$\mathcal{B}_{ab} = \text{curl } \sigma_{ab} - \frac{1}{2} \hat{\mathcal{D}}_{\langle a} \varpi_{b \rangle}^* \quad (6.35)$$

where the curl of a tensor is defined by

$$\text{curl } \sigma_{ab} \equiv I \gamma_{\langle a} \wedge u \wedge \hat{\mathcal{D}} \wedge \sigma_{b \rangle} = \epsilon_{cd \langle a} \hat{\mathcal{D}}^c \sigma_{b \rangle}^d \quad (6.36)$$

and the vorticity vector  $\varpi^*$  is

$$\varpi_a^* \equiv I \gamma_a \wedge u \wedge \varpi = \text{curl } u_a. \quad (6.37)$$

The curl of a PSTF tensor is also PSTF.

Further relations can be obtained from the Bianchi identity

$$\mathcal{D} \wedge \mathcal{R}_{ab} = 0. \quad (6.38)$$

Decomposing the Riemann tensor, using Einstein's equation and the symmetries of the Weyl tensor this becomes

$$\mathcal{D} \cdot \mathcal{W}_{ab} = \kappa \mathcal{D}_{[a} \mathcal{I}_{b]} - \frac{1}{6} \kappa \mathcal{D} \mathcal{T} \cdot (\gamma_a \wedge \gamma_b). \quad (6.39)$$

Contracting with  $u^b$  and using the energy conservation equations gives a somewhat lengthy expression for the right hand side. For the left hand side we use

$$u^b \mathcal{D} \cdot \mathcal{W}_{ab} = \hat{\mathcal{D}} \cdot \mathcal{E}_a u - \dot{\mathcal{E}}_a - \theta \mathcal{E}_a - I u \wedge \hat{\mathcal{D}} \wedge \mathcal{B}_a \quad (6.40)$$

which follows from writing

$$u^b \mathcal{W}_{ab} = (\mathcal{E}_a + I \mathcal{B}_a) u. \quad (6.41)$$

Projecting the resulting equation parallel to  $u$  gives the constraint equation

$$\hat{\mathcal{D}} \cdot \mathcal{E}_a = \frac{1}{2} \kappa (\hat{\mathcal{D}} \cdot \pi_a + \frac{2}{3} \theta q_a + \frac{2}{3} \hat{\mathcal{D}}_a \rho) \quad (6.42)$$

and the PSTF part gives a propagation equation for  $\mathcal{E}_a$ :

$$\dot{\mathcal{E}}_{ab} + \theta \mathcal{E}_{ab} - \text{curl } \mathcal{B}_{ab} = \frac{1}{2} \kappa (\dot{\pi}_{ab} + \frac{1}{3} \theta \pi_{ab} - [\rho + p] \sigma_{ab} - \hat{\mathcal{D}}_{\langle a} q_{b \rangle}). \quad (6.43)$$

To obtain similar equations for  $\mathcal{B}_a$  consider the dual of the Weyl tensor  $\mathcal{W}_{ab}^*$  (which is equal to  ${}^* \mathcal{W}_{ab}$  from the symmetry of the Weyl tensor). The Bianchi identity implies

$$u^b \mathcal{D} \cdot \mathcal{W}_{ab}^* = -I u \wedge \hat{\mathcal{D}} \wedge \mathcal{E}_a - u \hat{\mathcal{D}} \cdot \mathcal{B}_a + \dot{\mathcal{B}}_a + \theta \mathcal{B}_a. \quad (6.44)$$

and the decomposition of the Riemann tensor gives

$$u^b \mathcal{D} \cdot \mathcal{W}_{ab}^* = \frac{1}{2} \kappa \gamma_a \wedge u \wedge \gamma^c \wedge \mathcal{D} \mathcal{I}_c + \frac{1}{6} \kappa I (\mathcal{D} \mathcal{T}) \wedge \gamma_a \wedge u. \quad (6.45)$$

Expanding the stress-energy tensor and equating terms gives two new equations. The projection parallel to  $u$  gives the constraint equation

$$\hat{\mathcal{D}} \cdot \mathcal{B}_a = \frac{1}{2} \kappa (\text{curl } q_a + [\rho + p] \varpi_a^*) \quad (6.46)$$

and the PSTF part gives the propagation equation

$$\dot{\mathcal{B}}_{ab} + \theta \mathcal{B}_{ab} + \text{curl } \mathcal{E}_{ab} = -\frac{1}{2} \kappa \text{curl } \pi_{ab}. \quad (6.47)$$

It is useful to note that  $\mathcal{E}_a$  and  $\mathcal{B}_a$  are independent of frame to first order—changing frame to  $u' = u + v$  where  $v$  is first order only gives second order changes to the decomposition of the Weyl tensor. Frames that reduce to that of fundamental observers in the FRW limit will be always be related by first order quantities, so the components of the Weyl tensor are the same in any frames that we shall consider. Since  $\pi_{ab}$  is frame invariant to first order it follows from (6.43) that the quantity

$$(\rho + p) \sigma_{ab} + \hat{\mathcal{D}}_{\langle a} q_{b \rangle} \quad (6.48)$$

is also frame independent, which is readily checked. These frame independent quantities are the covariant equivalent of the gauge invariant variables (up to linear combinations) frequently used in other texts (see [11] for a good review).

## 6.4 The spatial curvature

The spatial 3-curvature is defined by the commutator of the projected derivatives. Acting on a projected vector  $v$  for example we have

$$[\hat{\mathcal{D}}_a, \hat{\mathcal{D}}_b]v = -{}^{(3)}\mathcal{R}_{ab} \cdot v \quad (6.49)$$

(the sign is conventional). This definition only applies in the absence of vorticity since, using the definition of the projected derivatives and the Ricci equation, we have

$$[\hat{\mathcal{D}}_a, \hat{\mathcal{D}}_b]v = h_a^c h_b^d \underline{H} \mathcal{R}_{cd} \cdot v + (\hat{\mathcal{D}}_a u \wedge \hat{\mathcal{D}}_b u) \cdot v - 2\varpi_{ab} \underline{H} \dot{v}. \quad (6.50)$$

When talking about the 3-curvature we implicitly assume that the vorticity vanishes (which will be true for scalar modes, as described in the next section). If  $u$  is orthogonal to hypersurfaces of constant  $t$ , we have  $u \propto \mathcal{D}t$  and the vorticity vanishes. In this case it clearly makes sense to talk about the curvature in the hypersurface.

Taking the trace of the 3-curvature and using Einstein's equation gives a relation for the 3-Ricci tensor

$${}^{(3)}\mathcal{R}_a = \mathcal{E}_a - \frac{1}{2} \kappa \pi_a + \frac{2}{3} \kappa \rho \underline{H}_a - \frac{1}{3} \theta \sigma_a - \frac{2}{9} \theta^2 \underline{H}_a \quad (6.51)$$

and contracting again gives the scalar 3-curvature

$${}^{(3)}\mathcal{R} = 2\kappa\rho - \frac{2}{3}\theta^2. \quad (6.52)$$



The first order variable  $\eta_a$  defined by the projected derivative of the 3-curvature scalar is given by

$$\eta_a \equiv \frac{1}{2}S\hat{\mathcal{D}}_a^{(3)}\mathcal{R} = \kappa\mathcal{X}_a - \frac{2}{3}\theta\mathcal{Z}_a. \quad (6.53)$$

We could also consider the spatial divergence of the 3-Ricci tensor, but using (6.33) and (6.42) we have  $S\hat{\mathcal{D}}_a^{(3)}\mathcal{R}_a = \eta_a$  which could equally well serve as the definition of  $\eta$ . The propagation equation for  $\eta_a$  can be obtained from (6.30) and (6.18) for the zero order variables by using (6.16) to commute the time and spatial derivatives giving

$$\dot{\eta}_a + \frac{2}{3}\theta\eta_a = -^{(3)}\mathcal{R}\dot{h}_a - S(\kappa\hat{\mathcal{D}}_a\hat{\mathcal{D}}\cdot q + \frac{2}{3}\theta\hat{\mathcal{D}}_a\hat{\mathcal{D}}\cdot A). \quad (6.54)$$

## 6.5 The background equations

In an exact FRW universe the 3-curvature will be proportional to  $S^{-2}$  as the spatial gradients will vary as  $1/S$ . We therefore define a constant  $K$  determining the spatial curvature in the exact FRW model

$$^{(3)}\mathcal{R}_{ab} \equiv \frac{K}{S^2}\underline{H}_a \wedge \underline{H}_b \quad ^{(3)}\mathcal{R} = \frac{6K}{S^2}. \quad (6.55)$$

Using this definition in equation (6.52) we obtain the Friedmann equation for the exact FRW universe

$$\frac{K}{S^2} = \frac{1}{3}\kappa\rho - \frac{1}{9}\theta^2. \quad (6.56)$$

The critical density  $\rho_{\text{crit}}$  is the energy density giving a flat universe ( $K = 0$ ). Writing  $\rho = \Omega\rho_{\text{crit}}$  we therefore have

$$\Omega = 1 + \frac{9K}{S^2\theta^2} \quad (6.57)$$

In the FRW limit the Raychaudhuri equation (6.30) becomes the other Friedmann equation

$$\dot{\theta} + \frac{1}{3}\theta^2 = -\frac{1}{2}\kappa(\rho + 3p). \quad (6.58)$$

These are all the equations we need to propagate the background model given values of the Hubble parameter  $\theta/3$  today,  $\Omega$  today ( $\Omega_{\text{tot}}$ ), and the equation of state of the various components present.

Combining the two Friedmann equations we have

$$\dot{\theta} = -\frac{3}{2}\kappa(\rho + p) + \frac{3K}{S^2}. \quad (6.59)$$

In a  $\Lambda$ -dominated universe, or during inflation, we have  $\rho \approx -p \approx \text{constant}$ . Neglecting the curvature term there is then a solution  $S \propto e^{\theta t/3}$  and we have exponential expansion. Since the energy density is roughly constant the  $1/S^2$  factor on the curvature term will rapidly become insignificant as the universe expands, justifying our neglect of the curvature. The value of  $\Omega$  is therefore driven to being very close to unity. So one of the key predictions of many models of inflation is that the universe today is close to the critical density. However it is possible to contrive models that are not flat, for example see [67] and references therein.

In order distinguish between the possibilities we shall need to take account of the curvature when calculating predictions for the non-flat models.

In a flat matter or radiation dominated universe the universe decelerates under the gravitational attraction of the matter. During matter or radiation domination  $\rho \propto S^{-3}$  or  $\rho \propto S^{-4}$ . Observations suggest  $\Omega$  is of order one today, so whilst the curvature may have been important in the late evolution of the universe, it can safely be neglected in the early radiation or matter dominated eras. Conditions on the last scattering surface will therefore be roughly independent of curvature. What we need to take account of is how those conditions map onto what we see today.

## 6.6 Scalar, Vector, Tensor decomposition

At this point it is convenient to write the first-order variables as sums of terms that can be written as derivatives of scalars, vectors or tensors. We then also refer to a quantity that can be derived from derivatives of a scalar as a Scalar. A Vector would then be something that cannot be derived from derivatives of a scalar but can be derived from derivatives of a vector. And a Tensor can only be derived from derivatives of a rank-two tensor. We do not need to consider higher than rank-two Tensors because there are no terms in the stress-energy tensor of rank higher than two and the higher rank equations therefore decouple.

Since the Scalar variables can be derived from a scalar field we can expand in terms of  $Q^k$ , the zero order eigenfunctions of the comoving Laplacian  $S^2\hat{\mathcal{D}}^2$ :

$$S^2\hat{\mathcal{D}}^2Q^k \equiv S^2\hat{\mathcal{D}}^a\hat{\mathcal{D}}_aQ^k = k^2Q^k. \quad (6.60)$$

where  $\dot{Q}_k = 0$  to zero order. Similarly the Tensor equations can be expanded in terms of PSTF tensor harmonics  $Q_{ab}^{Tk}$  where

$$S^2\hat{\mathcal{D}}^2Q_{ab}^{Tk} = k^2Q_{ab}^{Tk} \quad (6.61)$$

and  $\hat{\mathcal{D}}^aQ_{ab}^{Tk} = 0$ . The tensor harmonics can be further split into harmonics with electric or magnetic parity.

The vector equations can be expanded in a similar way, however the vector modes decay and are unimportant in the evolution of a standard universe (e.g. one without topological defects).

## 6.7 The scalar equations

For scalar modes the vorticity is second order, from (6.13). All the curl terms also vanish, so from (6.35) we see that the magnetic part of the Weyl tensor  $\mathcal{B}_a$  also vanishes for scalar modes. We expand the other variables in terms of  $Q_{A_l}^k$ , PSTF tensors derived recursively from the scalar harmonics  $Q^k$  via

$$Q_{A_l}^k = \frac{S}{k}\hat{\mathcal{D}}_{\langle a_l}Q_{A_{l-1}}^k. \quad (6.62)$$

Our equations contain terms with the divergence of PSTF quantities, so to expand in terms of the  $Q_{A_l}^k$  we need to know the divergence of  $Q_{A_l}^k$ . This will be another PSTF tensor proportional to  $Q_{A_{l-1}}^k$  so that  $\hat{\mathcal{D}}^a Q_{aA_{l-1}}^k = x_l Q_{A_{l-1}}^k$ . In the FRW limit the commutator of two spatial derivatives on a tensor  $\mathcal{V}_c$  gives

$$[\hat{\mathcal{D}}_a, \hat{\mathcal{D}}_b] \mathcal{V}_c = \frac{K}{S^2} 2\mathcal{V}_{[a} \gamma_{b]} \cdot \gamma_c \quad (6.63)$$

which can be shown from (6.49). Extending this to an  $l$ -rank tensor and using (6.62) we find

$$\hat{\mathcal{D}}^2 Q_{A_l}^k = \frac{k^2}{S^2} [1 - l(l+1)\frac{K}{k^2}] Q_{A_l}^k. \quad (6.64)$$

Using (6.6) we then have

$$\hat{\mathcal{D}}^a Q_{aA_{l-1}}^k = x_l Q_{A_{l-1}}^k = \frac{1}{l} \frac{k}{S} [1 - 2l(l-1)\frac{K}{k^2}] Q_{A_{l-1}}^k + \frac{(2l-3)(l-1)}{l(2l-1)} x_{l-1} Q_{A_{l-1}}^k. \quad (6.65)$$

The solution to this recurrence relation for  $x_l$  gives

$$\hat{\mathcal{D}}^{a_l} Q_{A_{l-1}a_l}^k = \frac{l}{2l-1} \frac{k}{S} \beta_l Q_{A_{l-1}}^k \quad (6.66)$$

where for  $l > 0$  we have defined

$$\beta_l \equiv 1 - (l^2 - 1)K/k^2. \quad (6.67)$$

We now define scalar coefficients for the harmonic expansion in terms of  $Q_{A_l}^k$  as follows

$$\begin{aligned} \mathcal{X}_a &= \sum_k k \mathcal{X} Q_a^k & q_a &= \sum_k q Q_a^k & \pi_{ab} &= \sum_k \Pi Q_{ab}^k \\ \mathcal{Z}_a &= \sum_k \frac{k^2}{S} \mathcal{Z} Q_a^k & \sigma_{ab} &= \sum_k \frac{k}{S} \sigma Q_{ab}^k & \eta_a &= \sum_k \frac{k^3}{S^2} \eta Q_a^k \\ h_a &= \sum_k k h Q_a^k & A_a &= \sum_k \frac{k}{S} A Q_a^k & \mathcal{E}_{ab} &= - \sum_k \frac{k^2}{S^2} \phi Q_{ab}^k \end{aligned} \quad (6.68)$$

where from now on the  $k$ -dependence is implicit. Factors of  $k$  are included to make the non-stress-energy expansion coefficients dimensionless. The minus sign in the definition of  $\phi$  is to make the sign convention agree with that of the Newtonian potential.

It is now convenient to use conformal time  $\eta$  where  $S d\eta = dt$ , and we now use dashes to denote derivatives with respect to conformal time. For example the conformal time expansion rate is  $H \equiv S'/S = S\theta/3$ .

The constraint equations derived from the Einstein equation are

$$\frac{2}{3} k^2 (\beta_2 \sigma - \mathcal{Z}) = \kappa S^2 q \quad (6.69)$$

$$k^3 \beta_2 \phi = -\frac{1}{2} \kappa S^2 [k(\beta_2 \Pi + \mathcal{X}) + 3Hq]. \quad (6.70)$$

and the propagation equations are

$$k(\sigma' + H\sigma) - k^2(\phi + A) = -\frac{1}{2} \kappa S^2 \Pi \quad (6.71)$$

$$k^2(\phi' + H\phi) = \frac{1}{2}\kappa S^2 [k(\rho + p)\sigma + kq - \Pi' - H\Pi] \quad (6.72)$$

The equations relating the other variables are

$$k^2\eta = \kappa S^2\mathcal{X} - 2kHZ \quad h' = \frac{1}{3}kZ - HA. \quad (6.73)$$

and the gradient of the 3-curvature evolves as

$$k\eta' = -2KZ - \kappa S^2q - 2\beta_2kHA. \quad (6.74)$$

Finally the momentum conservation equation is

$$q' + 4Hq + (\rho + p)kA - k\mathcal{X}^p + \frac{2}{3}k\beta_2\Pi = 0 \quad (6.75)$$

and the propagation equation for the density perturbation is

$$\mathcal{X}' + 3h'(\rho + p) + 3H(\mathcal{X} + \mathcal{X}^p) + kq = 0 \quad (6.76)$$

## 6.8 Evolution of the potential

During single field inflation the anisotropic stress is exactly zero, and during the later evolution of the universe it is often small. In this section we review some well known results for the evolution of scalar perturbations in the absence of anisotropic stress [11].

Taking the time derivative of (6.72), setting  $\Pi = 0$  and using (6.75), (6.71) and (6.70) we obtain the single equation

$$\phi'' + 3H(1 + \frac{p'}{\rho'})\phi' + k^2\beta_2\frac{p'}{\rho'}\phi + \left[(1 + 3\frac{p'}{\rho'})H^2 + 2H' - K\right]\phi = \frac{1}{2}\kappa S^2(\mathcal{X}^p - \frac{p'}{\rho'}\mathcal{X}). \quad (6.77)$$

For adiabatic perturbations the right hand side vanishes. On large scales ( $k \ll H$ ) we can neglect terms in  $k^2$  and for adiabatic perturbations this becomes

$$\phi'' + 3H(1 + \frac{p'}{\rho'})\phi' + \left[2H' + (H^2 - K)(1 + 3\frac{p'}{\rho'})\right]\phi = 0. \quad (6.78)$$

Defining the zero order variable

$$B^2 \equiv H^2 - H' + K = \frac{1}{2}\kappa S^2(\rho + p) \quad (6.79)$$

and using

$$\frac{p'}{\rho'} = -\frac{1}{3} - \frac{2B'}{3BH} \quad (6.80)$$

we obtain

$$\phi'' + 2\frac{B}{S}\left(\frac{S}{B}\right)'\phi' + \left[2B\left(\frac{H}{B}\right)' + \frac{2KB'}{BH}\right]\phi = 0. \quad (6.81)$$

We can remove the term in  $\phi'$  by changing variables to  $u = S\phi/B$  and after a little work the equation becomes simply

$$u''/u = \Theta''/\Theta \quad (6.82)$$

where  $\Theta = H/SB$ . There is therefore a decaying solution where  $u_d = \Theta$  and hence  $\phi_d = H/S^2$ . As there is no first derivative term the Wronskian will be a constant and the full solution is

$$u = C_1 \Theta + C_2 \Theta \int \frac{d\eta}{\Theta^2}. \quad (6.83)$$

Using the decaying solution we can write the Wronskian in terms of  $\phi$  for the other independent solution to derive the conserved quantity

$$\chi = \phi + \frac{2}{3}\Omega^{-1} \frac{H^{-1}\phi' + \phi}{1+w} + \frac{2}{3} \frac{\Omega^{-1} - 1}{1+w} \phi \quad (6.84)$$

where  $w = p/\rho$  and  $\Omega = 1 + K/H^2$ . This can be related to the curvature perturbation via

$$\bar{\eta} = -2\beta_2 \left[ \phi + \frac{2}{3}\Omega^{-1} \frac{H^{-1}\phi' + \phi}{1+w} \right] \quad (6.85)$$

where  $\bar{\eta}$  is the curvature perturbation evaluated in the frame of the total energy (where  $q = 0$ ). The large scale curvature perturbation  $\bar{\eta}$  is therefore conserved in a flat universe. At early times the expansion rate  $H$  is large so  $\Omega$  will be very close to one. Hence  $\bar{\eta}$  will be accurately conserved on super-Hubble scales in the early universe for all geometries compatible with  $\Omega \approx 1$  today (assuming negligible anisotropic stress).

In a flat universe when the equation of state is constant ( $\dot{w} = 0$ ) we have  $p'/\rho' = w$ . In this case the coefficient of  $\phi$  in (6.78) is  $H^2(1 + 3p'/\rho') + 2H' = 0$  and it follows that there is a solution  $\phi = \text{const}$ . The ‘growing’ mode solution is therefore a constant and so

$$\chi = \frac{5 + 3w}{3(1+w)} \phi. \quad (6.86)$$

This is the useful result that on large scales the potentials are constant when there is no significant curvature and the equation of state is constant. In particular the large scale potentials will be constant during matter domination ( $w = 0$ ) and radiation domination ( $w = 1/3$ ), and the amplitudes in the two regimes will be related by a factor of 9/10. This is of course not an exact result since we have ignored the effects of anisotropic stress which may be significant during the transition from radiation to matter domination.

## 6.9 Perturbations in single-field inflation

Deriving propagation equations for fluctuations of a scalar field during inflation is straightforward and simply amounts to inserting the correct form for the terms in the decomposition of the stress-energy tensor into the perturbation equations. Covariant equations for inflationary perturbations have been derived before using a particular choice of frame [68]. Here we give results valid in any frame.

We consider a single scalar field with action

$$S_\psi = \int dV_4 \left[ \frac{1}{2} \mathcal{D}\psi \cdot \mathcal{D}\psi - V(\psi) \right] \quad (6.87)$$

which yields the field equation

$$\mathcal{D} \cdot \mathcal{D}\psi = -V_{,\psi} \quad (6.88)$$

and stress-energy tensor

$$\mathcal{T}_{ab} = \mathcal{D}_a\psi\mathcal{D}_b\psi + (V - \frac{1}{2}\mathcal{D}\psi \cdot \mathcal{D}\psi)\gamma_a \cdot \gamma_b. \quad (6.89)$$

For the homogeneous and isotropic background solution the only non-zero terms are

$$\rho = \frac{1}{2}\dot{\psi}^2 + V \quad \text{and} \quad p = \frac{1}{2}\dot{\psi}^2 - V \quad (6.90)$$

The Friedmann equations combine to give the background equation

$$\frac{1}{2}\kappa\psi'^2 = H^2 - H' + K \quad (6.91)$$

and hence  $B^2 = \frac{1}{2}\kappa\psi'^2$ . The background field equation is

$$\psi'' + 2H\psi' + S^2V_{,\psi} = 0. \quad (6.92)$$

We consider linear perturbations about the background and characterize perturbations of the scalar field by the first order covariant vector

$$\mathcal{V}_a = S\hat{\mathcal{D}}_a\psi. \quad (6.93)$$

The other components of the stress-energy tensor are then

$$S^2q_a = \psi'\mathcal{V}_a \quad \pi_{ab} = 0 \quad (6.94)$$

and  $\mathcal{X}_a = S\hat{\mathcal{D}}_a\rho$  is given by

$$S^2\mathcal{X}_a = \psi'\mathcal{V}'_a + SA_a\psi'^2 + \mathcal{V}_aS^2V_{,\psi}. \quad (6.95)$$

Taking the spatial derivative of the field equation and commuting derivatives gives

$$\mathcal{V}''_a + 2H\mathcal{V}'_a + S\psi'(A'_a + 2HA_a) + 2SA_a\psi'' + S\mathcal{Z}_a\psi' + S\hat{\mathcal{D}}_a\hat{\mathcal{D}} \cdot \mathcal{V} = -S^2\mathcal{V}_aV_{,\psi\psi} \quad (6.96)$$

though this does not appear to be very useful.

We now obtain the scalar equations by expanding in terms of  $Q^k$  where

$$\mathcal{V}_a = \sum_k k\mathcal{V}Q_a^k. \quad (6.97)$$

The propagation equation for the potential becomes

$$\phi' + H\phi = \frac{1}{2}\kappa\psi' \left( \mathcal{V} + \psi' \frac{\sigma}{k} \right) \quad (6.98)$$

and the constraint equation is

$$-k^2\beta_2\phi = \frac{1}{2}\kappa \left[ \psi'\mathcal{V}' + A\psi'^2 + S^2\mathcal{V}V_{,\psi} + 3H\psi'\mathcal{V} \right]. \quad (6.99)$$

The variable

$$\tilde{\mathcal{V}} \equiv \mathcal{V} + \psi' \frac{\sigma}{k} \quad (6.100)$$

is frame independent, from (6.48). In the zero shear frame (the Newtonian gauge) we have  $\tilde{\mathcal{V}} = \mathcal{V}$  and it also follows from equation (6.71) that  $A = -\phi$ . Writing the equations in terms of frame invariant variables in the  $\sigma = 0$  frame we have

$$\phi' + H\phi = \frac{1}{2}\kappa\psi'\tilde{\mathcal{V}} \quad (6.101)$$

and

$$-k^2\beta_2\phi = \frac{1}{2}\kappa \left[ \psi'\tilde{\mathcal{V}}' - \phi\psi'^2 + S^2\tilde{\mathcal{V}}V_{,\psi} + 3H\psi'\tilde{\mathcal{V}} \right]. \quad (6.102)$$

Since these the equations contain only frame independent variables they must apply in all frames. They are equivalent to the well known gauge invariant equations, for example see [11].

Since the anisotropic stress is zero equation (6.77) applies where the term on the right hand side is determined by

$$\frac{1}{2}\kappa S^2 \left( \mathcal{X}^p - \frac{p'}{\rho'} \mathcal{X} \right) = -\frac{4}{3}k^2\beta_2 \left[ 2 + \frac{\psi''}{H\psi'} \right] \phi = \frac{4}{3}k^2\phi \frac{S^2 V_{,\phi}}{H\psi'}. \quad (6.103)$$

On large scales where we neglect  $k^2$  terms this is zero if  $K = 0$ , but not otherwise (because of the  $\beta_2$  factor). The quantity  $\chi$  is therefore not conserved in a non-flat universe because we have an additional curvature factor in the equation. However in a flat universe, or for a model in which the curvature term is negligible, it is conserved. Assuming that the perturbations remain adiabatic and the large scale anisotropic stresses remain negligible we can then use the constancy of  $\chi$  to relate super-Hubble perturbations in the scalar field to perturbations after re-heating.

In general we have

$$\phi'' + 2 \left( H - \frac{\psi''}{\psi'} \right) \phi' + 2 \left( H' - H \frac{\psi''}{\psi'} \right) \phi + (\beta_2 k^2 - K) \phi = 0. \quad (6.104)$$

which gives the general propagation equation for  $u \equiv S\phi/\psi'$  (differing by a constant from before, for later convenience) as

$$u'' + (k^2 - \Theta''/\Theta) u = 2K \left( 2 + \frac{\psi''}{H\psi'} \right) u. \quad (6.105)$$

## 6.10 Quantum generation of perturbations

**note this section is basically wrong.**

One of the successes of the inflationary scenario is that it can explain the generation of perturbations. We briefly review the argument here in order to see the form that we might expect the perturbations to take if inflation is correct. For further details see [11, 69].

We assume that during inflation the expansion is approximately exponential. After a few e-foldings any initial distribution of particles will have been diluted away, and on scales

much smaller than the Hubble radius we can assume that we are in a vacuum state of Minkowski space.

Quantum mechanics says that the inflaton fields will obey the equal time commutation relations

$$[\psi(x_1), \psi'(x_2)] = i\delta^3(x_1 - x_2). \quad (6.106)$$

We expand the field in terms of creation and annihilation operators as

$$\psi(x) = \psi_0 + \frac{1}{\sqrt{2}} \int \frac{d^3k}{(2\pi)^{2/3}} \left( \mathcal{V}_k^* e^{ikx} a_k^- + \mathcal{V}_k e^{-ikx} a_k^+ \right) \quad (6.107)$$

where

$$[a_k^-, a_{k'}^+] = \delta^3(k - k'). \quad (6.108)$$

For large  $k$  the perturbation variable obeys the wave equation

$$\mathcal{V}'' + k^2 \mathcal{V} = 0 \quad (6.109)$$

so that a positive frequency solution at time  $\eta_i$  is given by

$$\mathcal{V}_i = k^{-1/2} e^{-ik(\eta - \eta_i)}. \quad (6.110)$$

The  $k^{-1/2}$  factor comes from imposing consistency of (6.108) and (6.106). This gives us our initial condition for  $\mathcal{V}$ . We can relate it to the initial condition for  $u$  using (6.98) so that

$$|u_i| = \frac{|\mathcal{V}_i|}{k} = \frac{1}{k^{3/2}}. \quad (6.111)$$

The evolution equation (6.105) can then be evolved until each mode is well outside the Hubble radius. Then calculating the curvature perturbation, which is conserved on super-Hubble scales, we can calculate the initial power spectrum after reheating.

Using the commutation relation the correlation function for the potential is given by

$$\langle 0 | \phi(x) \phi(x+r) | 0 \rangle = \int \frac{dk}{k} \frac{\sin kr}{kr} |\delta_\phi|^2 \quad (6.112)$$

where

$$|\delta_\phi|^2 = \frac{\psi'}{S^2 (2\pi)^2} k^3 |u_k|^2 \quad (6.113)$$

is the power spectrum for  $\phi$ . In the initial state the factors of  $k$  cancel. Many inflation models therefore predict a nearly scale-invariant spectrum, with corrections arising from the slightly different time evolution of the different modes. To compute exact results we would need to know the exact form of the potential over the period in which relevant length scales leave the horizon.



## 6.11 PSTF expansion of the distribution function

After reheating the universe contains a mixture of components, photons, neutrinos, baryons, CDM, etc. We now need to work out how these evolve given an initial power spectrum for the potential when the modes are still far outside the horizon.

To study the evolution of the perturbations we need to analyse the propagation of the distribution function for the various species present. The distribution function will depend on momentum and position, and therefore depends on the direction of the momentum in the space orthogonal to  $u$ . In this section we discuss how to generate an irreducible decomposition of a general function of direction in terms of PSTF tensors, the covariant equivalent of a spherical harmonic expansion.

Consider an arbitrary function  $f(e)$  of direction in the 3-space orthogonal to  $u$  ( $e$  is a unit projected vector with  $e^2 = -1$ ). The function could be expanded in spherical harmonics  $Y_{lm}(\theta, \phi)$  of the angle of the vector  $e$ . However rank- $l$  PSTF tensors have  $2l + 1$  degrees of freedom and form an equivalent representation of the 3d rotation group. We can therefore equally well expand  $f(e)$  in terms of PSTF tensors:

$$f = \sum_{l=0}^{\infty} F_{A_l} e^{A_l} = F + F_a e^a + F_{ab} e^a e^b + \dots \quad (6.114)$$

where the  $F_{A_l}$  are PSTF. This is the covariant equivalent to a spherical harmonic expansion so

$$\sum_{m=-l}^l a_{lm} Y_{lm}(\theta, \phi) = F_{A_l} e^{A_l}. \quad (6.115)$$

The  $A_l$  notation is useful to represent a string of  $l$  indices, and the  $e^{A_l} \equiv e^{a_1} e^{a_2} \dots e^{a_l}$ . The reason we use a PSTF expansion is that the tensors in the expansion can be directly related to other physical quantities like the tensors that make up the decomposition of the stress-energy tensor.

Since the spherical harmonics are orthogonal we expect

$$F_{A_l} G_{B_n} \int e^{A_l} e^{B_n} d\Omega \propto \delta_{ln} F_{A_l} G^{A_l}. \quad (6.116)$$

Equivalently we can write the orthogonality relation

$$\int e_{\langle A_l \rangle} e^{\langle B_n \rangle} d\Omega = \delta_{ln} \Delta_l h_{\langle A_l \rangle}^{\langle B_l \rangle} \quad (6.117)$$

where  $h_{\langle A_l \rangle}^{\langle B_l \rangle} = h_{\langle a_1 \dots a_l \rangle}^{\langle b_1 \dots b_l \rangle}$  and the constant  $\Delta_l$  is to be determined. Evaluating the constant takes a little work but is essential to get obtain the correct numerical factors.

Integrated over angles the  $e^{A_l}$  must give an isotropic symmetric tensor and we have

$$\int e^{(A_{2l})} d\Omega = x_l h^{(A_{2l})} \quad (6.118)$$

where the  $x_l$  is to be determined. Contracting with  $h_{(A_{2l})}$  the term inside the integral becomes  $(-1)^l$ , and

$$h_{(A_{2l})}h^{(A_{2l})} = 2l + 1 \quad (6.119)$$

so we have  $x_l = (-1)^l 4\pi/(2l + 1)$  and hence

$$\int e^{(A_{2l})} d\Omega = \frac{(-1)^l}{2l + 1} h^{(A_{2l})}. \quad (6.120)$$

Using (6.6) we have

$$h_{\langle a_l}^{(A_l)} h_{A_{l-1}}^{B_{l-1}\rangle} = \frac{2l+1}{2l-1} h_{\langle A_{l-1}}^{B_{l-1}\rangle} \implies h_{\langle A_l}^{(A_l)} = 2l + 1 \quad (6.121)$$

so contracting with  $h_{\langle B_l}^{(A_l)} \rangle$  we have

$$h_{\langle B_l}^{(A_l)} \int e_{A_l} e^{B_l} d\Omega = (2l + 1) \Delta_l. \quad (6.122)$$

Using (6.120) we then have

$$\frac{4\pi(-1)^l}{2l + 1} h_{\langle B_l}^{(A_l)} h_{(A_l}^{B_l)} = (2l + 1) \Delta_l. \quad (6.123)$$

There are  $(2l)!$  index permutations in  $h_{(A_l}^{B_l)}$ , and the contribution to the contraction will come for those in which  $a$ 's and  $b$ 's alternate. There are  $(l!)^2$  ways to choose  $(a_i, b_j)$  pairs and  $2^l$  permutations within the pairs. We therefore have

$$\Delta_l = \frac{4\pi(-1)^l}{(2l + 1)^2} \frac{2^l (l!)^2}{(2l)!} h_{\langle B_l}^{(A_l)} h_{(A_l}^{B_l)} = \frac{4\pi(-2)^l (l!)^2}{(2l + 1)!}. \quad (6.124)$$

Further properties of PSTF tensors are given in [59].

## 6.12 The Boltzmann equation

The Boltzmann equation describes evolution of a distribution function  $f = f(x, p)$  along a path described by an affine parameter  $\tau$

$$\partial_\tau f = \mathcal{C} \quad (6.125)$$

where  $\mathcal{C}$  is a collision term. We define the affine parameter so that  $\partial_\tau x = p$  and decompose the four momentum  $p$  with respect to our reference velocity  $u$  as

$$p = Eu + \lambda e \quad (6.126)$$

where  $e^2 = -1$ . The stress energy tensor is obtained by integrating the distribution function over energy and angles

$$\mathcal{T}_{ab} = \int \frac{d^3(\lambda e)}{E} f(p) p_a p_b = \int dE d\Omega \lambda f(p) p_a p_b. \quad (6.127)$$

The  $e$  dependence of the distribution function  $f = f(x, \lambda, E, e)$  and the collision term can be expanded in terms of irreducible components as

$$f = \sum_{l=0}^{\infty} F_{A_l} e^{A_l} \quad \text{and} \quad \mathcal{C} = \sum_{l=0}^{\infty} C_{A_l} e^{A_l} \quad (6.128)$$

where the  $F_{A_l}$  and  $C_{A_l}$  are PSTF. Using this multipole expansion we can write the evolution of the distribution function as

$$\partial_\tau f = \sum_{l=0}^{\infty} [\partial_E F_{A_l} \partial_\tau E e^{A_l} + p \cdot \mathcal{D} F_{A_l} e^{A_l} + l F_{A_l} p \cdot \mathcal{D} e^{A_l} e^{A_{l-1}}]. \quad (6.129)$$

The component of the geodesic equation  $p \cdot \mathcal{D} p = 0$  in the  $e$  direction gives

$$\partial_\tau \lambda = E^2 e \cdot A + E \lambda (\sigma_{ab} e^a e^b - \frac{1}{3} \theta) \quad (6.130)$$

which then implies

$$h_{ab} p \cdot \mathcal{D} e^b = -\frac{E^2}{\lambda} (A_a + e \cdot A e_a) - E (e^b \varpi_{ba} + \sigma_{bc} e^b e^c e_a + \sigma_{ab} e^b) \quad (6.131)$$

which is first order. Linearizing we therefore have

$$\begin{aligned} \partial_\tau f &= \partial_E F (E \lambda A_a e^a + \lambda^2 \sigma_{ab} e^a e^b) \\ &+ \sum_{l=0}^{\infty} \left[ (E \dot{F}_{A_l} - \frac{1}{3} \lambda^2 \theta \partial_E F_{A_l}) e^{A_l} + \lambda \hat{\mathcal{D}}_a F_{A_l} e^a e^{A_l} \right] \end{aligned} \quad (6.132)$$

Following [10] we integrate over energy defining the energy integrated multipoles

$$J_{A_l}^{(i)} \equiv \Delta_l \int_m^\infty \lambda E^2 dE F_{A_l} \left( \frac{\lambda}{E} \right)^n \quad K_{A_l}^{(i)} \equiv \Delta_l \int_m^\infty \lambda E dE C_{A_l} \left( \frac{\lambda}{E} \right)^n \quad (6.133)$$

where  $n = l + 2i$  is the velocity weight. The factor  $\Delta_l$  is introduced so that the terms in the stress-energy tensor are simply

$$\rho = J^{(0)} \quad q_a = J_a^{(0)} \quad \pi_{ab} = J_{ab}^{(0)} \quad p = \frac{1}{3} J^{(1)} \quad (6.134)$$

which follows from (6.117) and (6.127), using the expansion of the distribution function and equating terms. It is also convenient to define

$$\rho^{(i)} = J^{(i)} \quad (6.135)$$

so that  $\rho = \rho^{(0)}$ . We can now multiply the Boltzmann equation by  $\lambda^{1+2i+l} E^{1-2i-l}$ , integrate over energy, and equate the corresponding PSTF terms giving

$$\begin{aligned} & \dot{J}_{A_l}^{(i)} + \frac{1}{3} \theta \left[ (1-n) J_{A_l}^{(i+1)} + (3+n) J_{A_l}^{(i)} \right] - \frac{l}{2l+1} \hat{\mathcal{D}}_{\langle a_l} J_{A_{l-1}}^{(i+1)} + \hat{\mathcal{D}}^a J_{a A_l}^{(i)} \\ & + \left[ (1+l+n) \rho^{(i+l-1)} + (3-l-n) \rho^{(i+l)} \right] (\delta_{1l} \frac{1}{3} A_{a_1} - \delta_{2l} \frac{2}{15} \sigma_{a_1 a_2}) = K_{A_l}^{(i)} \end{aligned} \quad (6.136)$$

for  $l > 0$  and the zero multipole equations

$$\dot{\rho}^{(i)} + \frac{1}{3}\theta \left[ (1-n)\rho^{(i+1)} + (3+n)\rho^{(i)} \right] + \hat{\mathcal{D}}^a J_a^{(i)} = K^{(i)}. \quad (6.137)$$

Here we have used (6.6). These are the propagation equations for the velocity-weighted multipoles and zero order densities. We characterize the perturbation to the  $\rho^{(i)}$  by the first order comoving spatial gradients

$$\chi_a^{(i)} \equiv S\hat{\mathcal{D}}_a \rho^{(i)}. \quad (6.138)$$

Taking the spatial gradient of the  $l = 0$  equations and commuting derivatives gives the propagation equation

$$\begin{aligned} \dot{\chi}_a^{(i)} + \left[ (1-n)\rho^{(i+1)} + (3+n)\rho^{(i)} \right] \dot{h}_a + S\hat{\mathcal{D}}_a \hat{\mathcal{D}}^b J_b^{(i)} \\ + \frac{1}{3}\theta \left[ (1-n)\chi_a^{(i+1)} + (3+n)\chi_a^{(i)} \right] = S\hat{\mathcal{D}}_a K^{(i)}. \end{aligned} \quad (6.139)$$

The above discussion is quite general, so we now move on to consider the scattering terms that will be important for propagation in the universe.

### 6.13 Thomson scattering

The only scattering terms we consider are those for the photons and electrons (which are tightly coupled to the baryons). After electron-positron annihilation the energy of photons will be much less than the mass of the electrons. The energy transfer during scattering can then be neglected and the Compton scattering formula we derived in Chapter 3 (equation (3.111)) reduces to the Thomson result

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{m^2} (\epsilon_f \cdot \epsilon_i)^2 \quad (6.140)$$

The only collision terms we shall consider are those for elastic scattering between non-relativistic electrons and photons since the  $1/m$  factor in the cross-section highly suppresses scattering from protons. Clearly the cross-section depends on the polarization, and to get accurate results we need to take account of this. The full treatment with polarization is rather more complicated and described in [9]. For simplicity we ignore polarization here and insert the correct polarization term by hand at the end.

For unpolarized radiation the cross-section for scattering from one electron at rest is

$$\sigma(\theta) = \frac{3}{16\pi} \sigma_T (1 + \cos^2 \theta) \quad (6.141)$$

where the total cross-section is  $\sigma_T$ . The amount that is scattered out of photon distribution  $f^\gamma(p)$  at a point is given by  $f_-^\gamma(p) = \sigma_T f^\gamma(p)$  in the rest frame. The amount gained is given by scattering into the phase space element

$$f_+^\gamma(p) = \frac{3}{16\pi} \sigma_T \int f^\gamma(p') [1 + (e^{(b)'} \cdot e^{(b)})^2] d\Omega' \quad (6.142)$$

so the total collision term is

$$\mathcal{C}^\gamma = n_e \sigma_T E^{(b)} [f_+^\gamma - f_-^\gamma] \quad (6.143)$$

where  $E^{(b)}$  is the energy in the electron-baryon frame. Integrating  $f_+^\gamma$  over energy in the baryon frame gives terms from the stress-energy tensor in the baryon frame:

$$\int dE^{(b)} E^{(b)3} f_+^\gamma = \frac{3}{16\pi} \left( \rho^{(b)} + e_a^{(b)} e_b^{(b)} \pi^{(b)ab} + p^{(b)} \right). \quad (6.144)$$

Photons are massless so  $E = \lambda$  and  $3p = \rho$ . If the baryons have a first order relative velocity  $v$  we have

$$u^{(b)} = \gamma_v (u + v) = u + v \quad (6.145)$$

to first order since  $\gamma_v = (1 + v^2)^{-1/2}$  is second order. The terms in the stress-energy tensor are therefore frame-independent and

$$\rho^{(b)} + e_a^{(b)} e_b^{(b)} \pi^{(b)ab} + p^{(b)} = \frac{4}{3} \rho + e^a e^b \pi_{ab} \quad (6.146)$$

to first order. The baryon velocity  $v$  is related to the baryon heat flux by  $q_b = (\rho_b + p_b)v$  (from  $q_b^{(b)} = 0$  and using  $u^{(b)} = u + v$ ).

The energy in the baryon frame is given by

$$E^{(b)} = E(1 + e \cdot v) \quad (6.147)$$

and

$$\int dE^{(b)} E^{(b)3} f_+^\gamma = (1 + e \cdot v)^3 \int dE E^2 E^{(b)} f_+^\gamma \quad (6.148)$$

so the energy integrated collision term is

$$\int dE E^2 \mathcal{C}^\gamma = \frac{3}{16\pi} n_e \sigma_T \left[ \frac{4}{3} (1 - 4e \cdot v) \rho_\gamma + e^a e \cdot \pi_a \right] - n_e \sigma_T \int dE E^3 f^\gamma(E, e). \quad (6.149)$$

Performing the multipole expansion of  $\mathcal{C}^\gamma$  and equating coefficients we have

$$K_a^\gamma = n_e \sigma_T \left( \frac{4}{3} \rho_\gamma v_a - q_a \right) \quad K_{ab}^\gamma = -\frac{9}{10} n_e \sigma_T \pi_{ab} \quad (6.150)$$

and for  $l > 2$  we have  $K_{A_l}^\gamma = -n_e \sigma_T J_{A_l}^\gamma$ . The baryon collision terms will just be the negative of these since the total stress-energy is conserved.

Including the polarization leads to  $\pi_{ab}$  being replaced by  $4\zeta_{ab}/3$  where

$$\zeta_{ab} \equiv \frac{3}{4} \pi_{ab} - \frac{9}{2} E_{ab}. \quad (6.151)$$

Here  $E_{ab}$  is the electric part of the linear polarization described in [9].

## 6.14 The scalar multipoles

We define scalar coefficients for the harmonic expansion of the multipole hierarchy in terms of  $Q_{A_l}^k$  as follows

$$J_{A_l}^{(i)} = \sum_k J_l^{(i)} Q_{A_l}^k \quad K_{A_l}^{(i)} = \sum_k K_l^{(i)} Q_{A_l}^k \quad \chi_a^{(i)} = \sum_k k \chi^{(i)} Q_a^k. \quad (6.152)$$

Inserting the harmonic expansion into the multipole equations and using (6.62) and (6.66) gives the  $l > 0$  propagation equations

$$J_l^{(i)'} + H \left[ (1 - 2i - l) J_l^{(i+1)} + (3 + 2i + l) J_l^{(i)} \right] + k \left\{ \frac{l+1}{2l+1} \beta_{l+1} J_{l+1}^{(i)} - \frac{l}{2l+1} J_{l-1}^{(i+1)} \right\} \\ + k \left[ (1 + 2l + 2i) \rho^{(i+l-1)} + (3 - 2l - 2i) \rho^{(i+l)} \right] (\delta_{1l} \frac{1}{3} A - \delta_{2l} \frac{2}{15} \sigma) = K_l^{(i)} \quad (6.153)$$

where  $J_0^{(i)} = \chi^{(i)}$ . Each spatial gradient evolves as

$$\chi^{(i)'} = -h' \left[ (1 - 2i) \rho^{(i+1)} + (3 + 2i) \rho^{(i)} \right] - H \left[ (1 - 2i) \chi^{(i+1)} + (3 + 2i) \chi^{(i)} \right] - k J_1^{(i)}. \quad (6.154)$$

The total matter variables are given by

$$\mathcal{X} = \sum \chi^{(0)} \quad q = \sum J_1 \quad \Pi = \sum J_2 \quad (6.155)$$

where the sum is over the different matter components present. We use a subscript on a scalar variable when we need to discuss a particular fluid component.

## 6.15 The fluid components

We shall assume that the fluid contains some baryons, cold dark matter, photons, massless neutrinos and/or massive neutrinos. The equations for photons and massless neutrinos simplify considerably since  $\lambda = E$ . It is useful to define

$$\rho_\gamma I_l \equiv J_{\gamma l} \quad \rho_\nu G_l \equiv J_{\nu l} \quad (6.156)$$

and the equations then simplify to

$$I_l' + k \left\{ \frac{l+1}{2l+1} \beta_{l+1} I_{l+1} - \frac{l}{2l+1} I_{l-1} \right\} = \\ -4h' \delta_l^0 - \frac{4}{3} k A \delta_l^1 + \frac{8}{15} k \sigma \delta_l^2 - S n_e \sigma_T (I_l - \delta_l^0 I_0 - \frac{4}{3} \delta_l^1 v - \frac{2}{15} \zeta \delta_l^2) \quad (6.157)$$

for the photons (in the absence of polarization  $\zeta = 3I_2/4$ ), and

$$G_l' + k \left\{ \frac{l+1}{2l+1} \beta_{l+1} G_{l+1} - \frac{l}{2l+1} G_{l-1} \right\} = -4h' \delta_l^0 - \frac{4}{3} k A \delta_l^1 + \frac{8}{15} k \sigma \delta_l^2 \quad (6.158)$$

for the massless neutrinos. Here  $I_0 = \chi_\gamma / \rho_\gamma$  and similarly for the neutrinos. The cold dark matter has no velocity dispersion by assumption, so from the definition of the  $J_{A_l}^{(i)}$  (equation

(6.133)) we only keep the  $l = 0, l = 1$  and  $i = 0$  terms. The  $l = 1$  term is necessary in general because the reference velocity  $u$  may not coincide with the CDM velocity. The resulting equations are

$$\Delta_c' = -3h' - kv_c \quad (6.159)$$

where  $\Delta_c = \chi_c/\rho_c$  and

$$v_c' + Hv_c + kA = 0 \quad (6.160)$$

where we have used  $q_c = \rho_c v_c$ . In the CDM gauge, in which  $v_c = 0$ , we see that the acceleration  $A = 0$ . We choose this gauge when we implement the equations numerically in the next chapter.

The baryons will be non-relativistic over the time scales of interest and so we only keep the  $l \leq 1, i \leq 1$  terms. The reason for keeping the pressure  $p_b = \frac{1}{3}\rho_b^{(1)}$  but not the  $l = 2$  term is that  $\pi_a$  is highly suppressed by scattering when there is a high electron density but the pressure could still be significant. At later times when the electron density is low the baryon momentum has been redshifted away and both will be insignificant. We only keep  $\hat{\mathcal{D}}p_b$  terms in the pressure which may be significant on small scales, dropping other terms, and assume that pressure variations are due to density variations so that

$$\hat{\mathcal{D}}p_b = c_s^2 \hat{\mathcal{D}}\rho_b. \quad (6.161)$$

The baryon equations are then

$$\Delta_b' = -3h' - 3c_s^2 H \Delta_b - kv \quad (6.162)$$

where  $\Delta_b = \chi_b/\rho_b$  and

$$J_{b1}' + 4HJ_{b1} - kc_s^2 \chi_b + kA = -Sn_e \sigma_T \rho_\gamma (\frac{4}{3}v - I_1). \quad (6.163)$$

For massive neutrinos we need to propagate the full energy-integrated equations. We discuss how to do this in practice later.

## 6.16 The scalar harmonics

Solutions for the  $Q^k$  eigenfunctions can be obtained explicitly in spherical polar coordinates by separation of variables:

$$(S^2 \hat{\mathcal{D}}^2 - k^2)Q^k = (S^2 \hat{\mathcal{D}}^2 - k^2)\Phi_l(\varrho)Y_{lm}(\theta, \phi) = 0 \quad (6.164)$$

where  $\varrho$  is the radial coordinate defined by the radial derivative of a scalar  $-Se \cdot \hat{\mathcal{D}} = \partial_\varrho$ . The  $Q^k$  are the zero order solutions, so we need to calculate  $\hat{\mathcal{D}}^2$  for an exact FRW universe. Isotropy restricts the 3-space to have a constant radius of curvature everywhere, and every 2-space cross section must also be isotropic. The surface of a sphere of radius  $r$  defines a closed isotropic 2-space. We can use  $\varrho$  to measure the geodesic distance from a point in the 2-space, and  $\theta$  as the angular coordinate about this point. The distance between two points at the same  $\varrho$  and separated by  $\theta$  will then be  $\theta r \sin_K(\varrho/r)$ . The quantity  $r \sin_K(\varrho/r)$

therefore defines the angular diameter distance. There are also isotropic hyperbolic spaces in which case the sine function is replaced by the sinh function. We use a  $K$  subscript on a function to denote that it should be of the hyperbolic variety for open models.

Moving up to an isotropic 3-space we can write the comoving spatial gradient as

$$Su \wedge \hat{\mathcal{D}} = \sigma_\varrho \partial_\varrho + \frac{\sigma_\theta}{r \sin_K(\varrho/r)} \partial_\theta + \frac{\sigma_\phi}{r \sin_K(\varrho/r) \sin \theta} \partial_\phi \quad (6.165)$$

where the  $\sigma$  are the corresponding frame vectors which obey

$$\partial_\theta \sigma_\varrho = \sigma_\theta \cos_K(\varrho/r) \quad \partial_\phi \sigma_\varrho = \sigma_\phi \sin \theta \cos_K(\varrho/r). \quad (6.166)$$

For small  $\varrho$  the result for  $\partial_\theta \sigma_\varrho$  differs by  $\sigma_\theta \varrho^2/2r^2$  from the flat result. The change of  $\sigma_\varrho$  due to the curvature going round a small loop at the origin is therefore given by  $A/r^2 \sigma_\theta$  where  $A$  is the area of the loop. Comparing with (6.55) for the 3-curvature we see that  $|K| = 1/r^2$ . In general we define  $r \equiv 1/\sqrt{|K|}$ .

Forming  $S^2 \hat{\mathcal{D}}^2$  using the above results we obtain the radial part of the harmonic equation

$$\frac{d^2 \Phi_l}{dx^2} + 2 \cot_K x \frac{d\Phi_l}{dx} + k^2 r^2 \Phi_l = \frac{l(l+1)}{\sin_K^2 x} \Phi_l \quad (6.167)$$

where  $x = \varrho/r$ . This equation can be written

$$\frac{1}{\sin_K x} \frac{d^2}{dx^2} (\sin_K x \Phi_l^\nu) = \left( \frac{l(l+1)}{\sin_K^2 x} - \nu^2 \right) \Phi_l^\nu. \quad (6.168)$$

where

$$\nu^2 \equiv \frac{k^2 + K}{|K|} = k^2 r^2 \pm 1. \quad (6.169)$$

In the flat case  $K = 0$  ( $\sin_K x = x$ ) the functions are simply spherical Bessel functions  $\Phi_l^\nu = j_l(kx)$ . In general solutions to this equation are called hyperspherical Bessel functions.

We want regular solutions at  $x = 0$  and normalize so that  $\Phi_0^\nu(0) = 1$  giving

$$\Phi_0^\nu(x) = \frac{\sin \nu x}{\nu \sin_K x}. \quad (6.170)$$

In open and flat models  $\nu^2$  can take any positive value. In closed models there are restrictions on the allowed values of  $\nu$  because the space is now spatially finite—for regularity at the point  $x = \pi$  we require  $\nu$  to be integer. For  $l \geq \nu$  the term in brackets on the RHS of (6.168) is always positive and there are no non-zero solutions that are zero at  $x = \pi$  as required for a regular solution. In closed models  $\nu$  is therefore restricted to be an integer with  $\nu > l$ . The  $\nu = 1$  solution gives  $\Phi_0^1 = \text{const}$  from (6.170), and therefore cannot describe perturbations. The  $\nu = 2$  mode can only contribute to  $l = 1$  modes since  $\beta > l$ , and hence only contributes to the CMB dipole.

It is straightforward to verify that functions satisfying the recursion relations

$$\frac{d\Phi_l^\nu}{dx} = l \cot_K x \Phi_l^\nu - kr \beta_{l+1} \Phi_{l+1}^\nu \quad (6.171)$$



and

$$(2l+1)\frac{d\Phi_l^\nu}{dx} = kr(l\Phi_{l-1}^\nu - (l+1)\beta_{l+1}\Phi_{l+1}^\nu). \quad (6.172)$$

will satisfy (6.168). These are the generalization of the usual spherical Bessel function recursion relations.<sup>1</sup> Differentiating equation (6.172)  $l-1$  times and evaluating at  $x=0$  we have the useful result

$$\frac{d^l}{dx^l}\Phi_l^\nu \Big|_{x=0} = \frac{(kr)^l l!}{(2l+1)!!} = (-kr)^l \frac{\Delta_l}{4\pi}. \quad (6.173)$$

## 6.17 The line of sight solution

It is possible to find a Green's function solution to (6.157). Physically this amounts to adding up the contributions to the anisotropy from each sphere down the past light cone from the observation point.

Equation (6.157) has a dependence on  $I_l$  on the right. We define the optical depth

$$\tau(\eta) = \int_\eta^{\eta_0} d\eta S n_e \sigma_T \quad (6.174)$$

so that  $S n_e \sigma_T = -\tau'$ . An integrating factor for (6.157) is then given by  $e^{-\tau}$  and the RHS then depends only on the lowest three multipoles.

From the recursion relation (6.172) with  $x = \eta/r$  we see that  $d^n \Phi_l^\nu / dx^n$  satisfies the LHS of (6.157) (taking the derivative of (6.172) gives the same equation for the derivative). Using (6.173) we have

$$\Phi_l^\nu|_{x=0} = \delta_l^0 \quad \frac{d\Phi_l^\nu}{dx} \Big|_{x=0} = \frac{1}{3}kr\delta_l^1 \quad \left[ \frac{1}{3}\Phi_l^\nu + \frac{1}{k^2 r^2} \frac{d^2\Phi_l^\nu}{dx^2} \right]_{x=0} = \frac{2}{15}\delta_l^2 \quad (6.175)$$

which we can use to construct a Green's function so that  $I_l$  today is given by

$$\begin{aligned} I_l|_{\eta_0} &= 4 \int_\eta^{\eta_0} d\eta e^{-\tau} \left\{ (k\sigma + \frac{1}{4}S n_e \sigma_T \zeta) \left[ \frac{1}{3}\Phi_l^\nu(\chi) + \frac{1}{k^2 r^2} \frac{d^2}{d\chi^2} \Phi_l^\nu(\chi) \right] \right. \\ &\quad \left. - (kA - S n_e \sigma_T v) \frac{1}{kr} \frac{d}{d\chi} \Phi_l^\nu(\chi) + \left[ \frac{1}{4}S n_e \sigma_T I_0 - h' \right] \Phi_l^\nu(\chi) \right\} \end{aligned} \quad (6.176)$$

where  $\chi = (\eta_0 - \eta)/r$ . This formula forms the basis of the efficient line of sight algorithm for computing CMB anisotropies [13] that we discuss in the next chapter.

---

<sup>1</sup>Note that our normalization of the  $\Phi_l^\nu$  differs by a curvature factor from that used by many other authors who use the functions  $\Phi_l^\nu \prod_{n=1}^l \beta_n^{1/2}$ . Our definition removes curvature factors and leads to a faster numerical implementation, though making the overall normalization dependent on curvature.

## 6.18 Instantaneous recombination approximation

The number density of electrons drops very sharply at recombination, so it is a reasonable approximation to assume that it is instantaneous (at least for scales larger than the depth of the last scattering surface). This is very useful for getting approximate analytic results and for understanding the basic physics of the structure of the CMB anisotropy spectrum, though on smaller scales the effects of variations during recombination can become important. With this approximation we have  $\tau'e^{-\tau} \approx \delta(\eta - \eta^*)$  where  $\eta^*$  is the conformal time of recombination. Gauge fixing to the CDM so that  $A = 0$  we have

$$I_l^0 = \left[ (I_0 + 4\frac{\sigma'}{k})\Phi_l^\nu - (v + \sigma)\frac{4}{kr}\frac{d}{dx}\Phi_l^\nu + 4\zeta \left( \frac{1}{3}\Phi_l^\nu + \frac{1}{k^2r^2}\frac{d^2\Phi_l^\nu}{dx^2} \right) \right]_{\eta^*} + 2 \int_{\eta^*}^{\eta_0} dt \dot{\phi} \Phi_l^\nu \quad (6.177)$$

for  $l > 2$  where we have used

$$\frac{\sigma''}{k} + \frac{1}{3}k(\sigma - \mathcal{Z}) = 2\phi' + A' - HA. \quad (6.178)$$

In the case that recombination is matter dominated the ‘‘monopole’’ term becomes

$$\frac{1}{4}I_0 + \frac{\sigma'}{k} = \frac{1}{4}\bar{\Delta}^\gamma + \frac{1}{3}\phi - \frac{2}{S^2\kappa\rho}H\phi' + \frac{2K}{\kappa\rho S^2}\phi. \quad (6.179)$$

The quantity  $\bar{\Delta}^\gamma = \Delta^\gamma + 4\frac{Hq}{k}$  is the photon perturbation in the total energy frame where  $q = 0$ . For flat models the last two terms contributing to the monopole are small, and the main contribution comes from the photon over-density and potential at last scattering. In the total energy frame the potential varies as mass/radius, and so varies approximately with the square of the perturbation radius. (This can be seen mathematically in equation (6.70) for small anisotropic stress). On large scales the Sachs-Wolfe term  $\frac{1}{3}\phi$  will therefore dominate. This represents the gravitational redshifting of the radiation as it climbs out a potential well. At smaller scales the anisotropies are dominated by the photon perturbation corresponding to hot and cold spots on the last scattering surface. Due to the sign difference between these two dominant terms we see that on large scales compressions correspond to cold spots in the CMB, but on small scales they correspond to hot spots.

The term proportional to the derivative of the Bessel function will be out of phase with the monopole term and determines the Doppler shifting due to the perturbation velocity. Contributions from this ‘dipole’ term dominate on scales that are in mid oscillation at last scattering, when the over-densities will be small but the velocities are high.

The integral of  $\dot{\phi}$  determines the integrated Sachs-Wolfe effect, the redshifting or blueshifting of photons as they climb in and out of potential wells on the way from the last scattering surface to our observation point. The potentials are small (compared to the density perturbations) on small scales, so this effect is only significant on large scales. For flat models we know from equation (6.86) that when the equation of state is a constant the large scale potentials will be constant. The integrated Sachs-Wolfe term will therefore be very small for flat models in which recombination is fully matter dominated. However in non-flat models, or in models where recombination happens before full matter domination, the potentials

will vary. There can also be a change in the equation of state if the cosmological constant comes to dominate, leading to an additional contribution.

The term depending on  $\zeta$  gives the small contribution from anisotropic stress and polarization at last scattering. The hyperspherical Bessel functions relate scales at a position on the line of sight light cone to angular scales observed on the sky today. Equivalent results in the Newtonian gauge (in which  $\sigma = 0$ ) are given in [70], along with an approximate analytic analysis of how the perturbations evolve before last scattering.

## 6.19 The CMB power spectrum

What we observe on the sky today are the CMB temperature anisotropies, which are usually quantified by the CMB power spectrum. If anisotropies originate from quantum fluctuations in inflation we would expect them to be Gaussian, in which case the power spectrum contains all the useful information. The Stefan-Boltzmann law relates the temperature to the radiation power, so that the average temperature is given by

$$T^4 \propto \frac{1}{4\pi} \int dE d\Omega E^3 f^\gamma(E, e) = \frac{1}{4\pi} \rho_\gamma. \quad (6.180)$$

The fractional temperature variation  $\delta_T(e)$  is then given by

$$[1 + \delta_T(e)]^4 = \frac{4\pi}{\rho_\gamma} \int dE E^3 f^\gamma(E, e) \quad (6.181)$$

and to first order

$$\delta_T(e) = \frac{4\pi}{4\rho_\gamma} \sum_{l=1}^{\infty} \Delta_l^{-1} I_{A_l} e^{A_l}. \quad (6.182)$$

We can also expand  $\delta_T(e)$  in spherical harmonics so that

$$\sum_{m=-l}^l a_{lm} Y_{lm}(\theta, \phi) = \frac{\pi}{\rho_\gamma} \Delta_l^{-1} I_{A_l} e^{A_l}. \quad (6.183)$$

Squaring the above expression and integrating over solid angles we have

$$\frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}|^2 = \frac{1}{4} \frac{\pi}{(\rho_\gamma)^2} \frac{(2l)!}{(-2)^l (l!)^2} I_{A_l} I^{A_l} \quad (6.184)$$

using (6.117). The CMB power spectrum can be described by  $C_l$  which is the statistical average of this equation

$$C_l \equiv \frac{1}{2l+1} \sum_{m=-l}^l \langle |a_{lm}|^2 \rangle. \quad (6.185)$$

For scalar modes the  $I_{A_l}$  are expanded in terms of  $Q_{A_l}^k$ . We therefore need to be able to evaluate  $Q_{A_l}^k Q^{k A_l}$  today where the  $Q_{A_l}^k$  are given by (from (6.62))

$$Q_{A_l}^k = \left(\frac{S}{k}\right)^l \hat{\mathcal{D}}_{\langle A_l} Q^k. \quad (6.186)$$

The scalar  $Q^k$  can be written as

$$Q^k = \Phi_l^\nu(x) \mathcal{Y}_{A_l}^{(lm)} e^{A_l} \quad (6.187)$$

where the  $\mathcal{Y}_{A_l}^{lm}$  are the PSTF representation of the spherical harmonics and  $e \cdot \hat{\mathcal{D}} \mathcal{Y}_{A_l}^{(lm)} = 0$ . Taking  $l$  derivatives in the  $e$  direction and evaluating at  $\chi = 0$  we have

$$e^{A_l} \hat{\mathcal{D}}_{A_l} Q^k \Big|_{\chi=0} = \frac{(-1)^l}{S^l} \frac{d^l}{d\chi^l} \Phi_L^\nu \Big|_{\chi=0} \mathcal{Y}_{A_L}^{Lm} e^{A_L}. \quad (6.188)$$

Using (6.173) and equating coefficients of  $e^{A_l}$  we have

$$Q_{A_l} \Big|_{\chi=0} = \frac{\Delta_l}{4\pi} \mathcal{Y}_{A_l}^{(lm)}. \quad (6.189)$$

From the orthogonality of the spherical harmonics we know

$$\int d\Omega \mathcal{Y}_{A_l}^{(lm)} \mathcal{Y}_{B_l}^{(lm')} e^{A_l} e^{B_l} = \delta_{mm'} \quad (6.190)$$

and so, using (6.117), we have

$$\mathcal{Y}_{A_l}^{(lm)} \mathcal{Y}^{(lm') A_l} = \delta_{mm'} \Delta_l^{-1}. \quad (6.191)$$

We can write the mode coefficients in terms of an  $l$ -dependent transfer function as

$$I_l = T_l(k) \phi_k \quad (6.192)$$

where the  $\phi_k$  are random variables with

$$\langle \phi_k \phi_{k'} \rangle = \mathcal{P}(k) \delta_{kk'}. \quad (6.193)$$

Here  $\mathcal{P}(k)$  is the power spectrum and  $\delta_{kk'}$  is defined so that  $\sum_k \delta_{kk'} \phi_k = \phi_{k'}$ . The  $T_l(k)$  are the transfer functions which determine the  $I_l$  for  $\phi_k = 1$ .

Putting all this together the power spectrum is given by

$$C_l = \frac{1}{16} \frac{1}{2l+1} \sum_k \mathcal{P}(k) [T_l(k)]^2. \quad (6.194)$$

We are free to choose how to define the sum over  $k$  modes for a given  $l$ . It still includes a sum over the  $2l+1$   $m$ -modes, but we can include factors of  $k$  as we please. The physical wavevector, the eigenvalue of the harmonic equation, is given by  $k_{\text{phys}} = k/S$ . We therefore choose the sum over  $k$  to be independent of scale factor, so

$$\sum_k = \sum_{lm} \int \frac{dk}{k} = \sum_{lm} \int \frac{dk_{\text{phys}}}{k_{\text{phys}}} \quad (6.195)$$

for open and flat models. For closed models the integral is replaced by a sum and we have

$$\int \frac{dk}{k} \rightarrow \sum_{\nu} \frac{\nu}{\nu^2 - 1} \quad (6.196)$$

using  $\nu^2 = k^2 r^2 + 1$ . We now have

$$C_l = \frac{1}{16} \int \frac{dk}{k} \mathcal{P}(k) [T_l(k)]^2 \quad (6.197)$$

where the integral is replaced by a sum for closed models.

For an adiabatic power spectrum generated from inflation we set the random variables  $\phi_k$  equal to the potentials. The initial power spectrum  $\mathcal{P}(k)$  is then approximately  $|\delta_\phi|^2$  evaluated when the mode has inflated well outside the horizon.

A scale invariant initial power spectrum is one in which the potential is independent of  $k$  and  $\mathcal{P}(k) = \text{const}$ . On large (super-Hubble) scales the potentials are constant in time and so a scale invariant spectrum implies scale invariant potentials at last scattering. On large scales the anisotropy is governed by the climb out of the potential wells at last scattering (assuming matter domination) and so  $I_l \propto \Phi_l^\nu$  from (6.179). The transfer functions are therefore approximately hyperspherical Bessel functions. For a flat models these become the usual Bessel functions  $j_l$  and we have

$$\int_0^\infty [j_l(x)]^2 \frac{dx}{x} = \frac{1}{2l(l+1)}. \quad (6.198)$$

Hence for flat models  $l(l+1)C_l$  will be independent of  $l$  (for small  $l$ ) for a scale invariant power spectrum. For open and closed models (and models that change equation of state significantly after last scattering) the potentials will vary after last scattering and the integrated Sachs-Wolfe effect causes a non-flat spectrum for  $l(l+1)C_l$  at low  $l$ .

This definition, where  $\mathcal{P}(k) = \text{const}$  for a scale invariant spectrum, is consistent with [9, 71] though differing from some other texts. From (6.85) we have  $\phi \propto \beta_2 \bar{\eta}$  for the super-Hubble modes ( $\phi' = w' = 0$  for super-Hubble modes during radiation domination). It follows that with our conventions the energy frame curvature perturbation has

$$\langle |\bar{\eta}_a|^2 \rangle \propto \int \frac{dk}{k} \frac{k^6}{S^4} \beta_2^2 \mathcal{P}(k) \quad (6.199)$$

from super-Hubble modes.

## 6.20 Initial conditions

It is possible to find power series solutions to the perturbation equations at early times. There are five independent regular solutions with vanishing neutrino anisotropic stress at time zero. The adiabatic growing mode is dominated by the potential perturbation. There are also three isocurvature modes in which the perturbation is dominated by the baryon, CDM and neutrino density perturbations, and a velocity mode dominated by the neutrino

heat flux. The initial conditions for these five modes are given in [72, 73]. The initial conditions for the massive neutrino hierarchies are discussed in the next section.

Adiabatic initial conditions are predicted by many models of inflation, including the single field example we have considered, and are often assumed to dominate. Less simple inflationary models can easily give more complicated predictions. The most general initial condition will be described by a  $5 \times 5$   $k$ -dependent symmetric matrix determining the relative amplitudes and correlations between the modes [74].

## 6.21 Massive neutrinos

The energy-integrated equations all reduce to one set of equations in the relativistic limit. Since the neutrinos will be highly relativistic in the very early universe we cannot use the energy integrated equations to propagate the neutrino hierarchy until the particles become non-relativistic because an infinite number of equations would be involved, all of approximately the same order. Instead we do not integrate the equations over energy, but propagate (an in principle infinite number) of hierarchies for different comoving momenta. In practice we only need to sample a fairly small number of momenta, and then integrate numerically over energy to give the terms in the energy integrated equations. This can be used to propagate the neutrino hierarchy as long as required, or it could be used to provide the starting conditions for the energy integrated equations once the neutrinos are no longer highly relativistic. The latter approach is much more efficient if the neutrinos are not very light, and similar schemes could be used to propagate perturbations for other forms of dark matter efficiently.

In terms of  $q \equiv S\lambda$  and  $\epsilon \equiv SE$  the geodesic equation (6.130) is

$$\partial_\tau q = \frac{1}{S}(\epsilon^2 A_a e^a + \epsilon q \sigma_{ab} e^a e^b + \frac{q^2}{S} e^a \hat{\mathcal{D}}_a S) \quad (6.200)$$

so the linearized Boltzmann equation gives

$$\partial_q F(\epsilon^2 A_a e^a + \epsilon q \sigma_{ab} e^a e^b + \frac{q^2}{S} e^a h_a) + \sum_{l=0}^{\infty} \left[ \epsilon \dot{F}_{A_l} e^{A_l} + q \hat{\mathcal{D}}_a F_{A_l} e^a e^{A_l} \right] = 0. \quad (6.201)$$

Using the CDM frame ( $A = 0$ ) and equating irreducible components this gives the  $l > 0$  multipole equations

$$\dot{F}_{A_l} - \frac{q}{\epsilon} \frac{l+1}{2l+3} \hat{\mathcal{D}}^c F_{cA_l} + \frac{q}{\epsilon} \hat{\mathcal{D}}_{\langle a} F_{A_{l-1}} \rangle + \delta_{l1} \frac{q}{S} h_a q \partial_q F + \delta_{l2} \sigma_{ab} q \partial_q F = 0 \quad (6.202)$$

and the zero multipole equation

$$\dot{F} = \frac{1}{3} \frac{q}{\epsilon} \hat{\mathcal{D}}^a F_a. \quad (6.203)$$

The  $l = 1$  equation contains the first order variable combination

$$\mathcal{V}_a \equiv S \hat{\mathcal{D}}_a F + h_a q \partial_q F \quad (6.204)$$

that we use to characterize perturbations to the distribution function. It integrates to give the comoving gradient of the density

$$\chi_a \equiv S \hat{\mathcal{D}}_a \rho_\nu = \frac{4\pi}{S^4} \int_0^\infty dq q^2 \epsilon \mathcal{V}_a. \quad (6.205)$$

Differentiating with respect to time and commuting derivatives gives the propagation equation

$$\dot{\mathcal{V}}_a = \frac{1}{3} \frac{q}{\epsilon} S \hat{\mathcal{D}}_a \hat{\mathcal{D}}^b F_b + \dot{h}_a q \partial_q F. \quad (6.206)$$

The scalar equations are obtained by expanding the variables in terms of the harmonics as

$$\mathcal{V}_a = \sum_k k F_0 Q_a^k \quad \text{and} \quad F_{A_l} = \frac{(2l+1)!}{(-2)^l (l!)^2} \sum_k F_l Q_{A_l}^k \quad (6.207)$$

giving the multipole equations

$$F_l' + k \frac{q}{\epsilon} \left\{ \frac{l+1}{2l+1} \beta_{l+1} F_{l+1} - \frac{l}{2l+1} F_{l-1} \right\} + \left[ \delta_{2l} \frac{2}{15} k \sigma - \delta_{1l} h' \right] q \partial_q F = 0. \quad (6.208)$$

We can calculate the energy integrated multipoles by integrating over  $q$

$$J_l^{(i)} = \frac{4\pi}{S^4} \int_0^\infty dq q^2 \epsilon \left( \frac{q}{\epsilon} \right)^{l+2i} F_l \quad (6.209)$$

and

$$\chi = \frac{4\pi}{S^4} \int_0^\infty dq q^2 \epsilon F_0. \quad (6.210)$$

Assuming the neutrinos are initially highly relativistic the initial conditions are determined by those for massless neutrinos as

$$F_0 = -\frac{G_0}{4} q \partial_q F \quad F_l = -\frac{G_l}{4} q \partial_q F. \quad (6.211)$$

The zero order distribution function is given by the Fermi-Dirac distribution

$$F(q) \propto \left[ \exp \frac{\sqrt{q^2 + m^2} S_d^2}{k T_d S_d} + 1 \right]^{-1} \quad (6.212)$$

where  $T_d$  and  $S_d$  are the temperature and scale factor at neutrino decoupling. We assume the particles will be highly relativistic at decoupling so the mass term can be dropped. The mass therefore only enters through the integrals over  $q$  and factors of  $\frac{q}{\epsilon}$  in the differential equations.

An efficient numerical integration of the hierarchy should propagate these equations whilst the neutrinos are relativistic, and then convert to the truncated integrated equations when non-relativistic. This is discussed in the next chapter. For a discussion of massive neutrinos and their effect on the CMB see [75].

## 6.22 The tensor equations

Tensor modes describe gravitational waves and can have a significant effect on the CMB anisotropies we observe. We shall not give the derivations of any of the tensor equations here. Whilst straightforward in principle, obtaining all the correct numerical factors takes considerable work and taking full account of the electric and magnetic tensor parts of the polarization is even more involved. For derivations and discussion see [8, 9].

Variables can be eliminated to reduce the covariant perturbation equations to two equations involving the harmonic expansion of the electric part of the Weyl tensor and the shear (harmonic coefficients  $E$  and  $\sigma$  respectively, defined analogously to the scalar coefficients). These have propagation equations

$$k^2(E' + HE) - \beta_0 k^3 \sigma + k \frac{1}{2} \kappa S^2 (\rho + p) \sigma = \frac{1}{2} \kappa S^2 (\Pi' + H\Pi) \quad (6.213)$$

and

$$k(\sigma' + H\sigma) + k^2 E = -\frac{1}{2} \kappa S^2 \Pi \quad (6.214)$$

where  $\beta_0 \equiv 1 + 3K/k^2$ . In the absence of anisotropic stress these combine to give

$$u'' + \left[ k^2 + 2K - \frac{(S^{-1})''}{S^{-1}} \right] u = 0 \quad (6.215)$$

where  $u \equiv S\sigma$  analogous to the scalar equation. On large scales (and neglecting the curvature) a decaying solution is therefore  $\sigma_d \propto 1/S^2$  and the conserved Wronskian is given by

$$k\chi \equiv \sigma' + 2H\sigma = H\sigma - kE. \quad (6.216)$$

As in the scalar case we can relate super-Hubble perturbations before and after reheating using the constancy of  $\chi$ . When the equation of state is a constant (like in the early radiation dominated era)  $E$  and  $H\sigma$  are both independently constant.

The variable<sup>2</sup>

$$\chi \equiv H \frac{\sigma}{k} - E + \frac{1}{2} \kappa \frac{S^2}{k^2} \Pi \quad (6.217)$$

satisfies  $\chi' = -k\beta_0\sigma$  and is proportional to the the variable  $H_T$  employed in non-covariant approaches [14]. In the absence of anisotropic stress  $\chi = \text{const}$  can be seen directly from the equation for  $u_T \equiv S\chi$

$$u_T'' + \left[ k^2 + 2K - \frac{S''}{S} \right] u_T = 0 \quad (6.218)$$

in the large-scale flat-space limit.

Tensor perturbations can be generated during inflation by quantum fluctuations. However giving a simple GTG-consistent derivation of the effect is not easy. Using GR it is found that an approximately scale invariant spectrum is expected, as in the scalar case [11]. After reheating anisotropic stresses can become significant and it is necessary to propagate the tensor modes of the Boltzmann hierarchy. The full set of equations is given in [9], and they are also listed in the appendix to the next chapter.

## 6.23 Conclusion

We have seen how covariant methods can be used to provide a physically transparent way of studying perturbations in cosmology. After a harmonic expansion we have arrived at a set of scalar equations that can be integrated numerically, as discussed in the next chapter.

---

<sup>2</sup>2005: Check this in CAMB notes



The scalar equations are equivalent to those derived by other methods, but the derivations here use only physically observable quantities and avoid any gauge ambiguities (except for the choice of frame, which is just specifying who is doing the observing). We have extended the covariant results of other workers to encompass massive neutrinos and the propagation of perturbations during single-field inflation.



## Chapter 7

# Efficient computation of CMB anisotropies

In this chapter we implement numerically the mode-expanded covariant perturbation equations to compute predictions for CMB anisotropies given initial conditions for the early radiation dominated era. The code uses the line of sight Green's function solution to compute the anisotropies efficiently for flat, open and closed geometries. We present new results for the polarization power spectra from scalar and tensor perturbations in closed models.

### 7.1 Introduction

The anisotropy of the CMB contains a great deal of information about the universe and therefore plays a key role in modern cosmology. Current data suggest that the universe is within a factor  $\Omega_{\text{tot}} \equiv \Omega_{\text{m}} + \Omega_{\Lambda} = 1 \pm 0.2$  of the critical density required for a flat geometry. However open ( $\Omega_{\text{tot}} < 1$ ) and closed models ( $\Omega_{\text{tot}} > 1$ ) still account for an important sector of the possible parameter space. Moreover, maximum likelihood searches require theoretical predictions over a much larger volume of parameter space to establish reliable error estimates on the parameters under consideration. It is therefore vital to have a fast and accurate method for calculating anisotropies for models at least within the range  $0.4 < \Omega_{\text{tot}} < 1.6$ .

The widely-used CMBFAST code [13, 14] uses the synchronous gauge equations to compute predictions for CMB anisotropies. However until very recently this did not support closed models, excluding an important section of parameter space. In this chapter we describe efficient code for computing CMB anisotropies in all geometries. Though CMBFAST does now support closed models the closed model code described here was developed independently and is considerably faster and more robust. It also provides a semi-independent check on the results of CMBFAST.

In the previous chapter we described the linearized equations governing the propagation of perturbations in nearly FRW models and the relation to the CMB anisotropies. We only considered the scalar anisotropies in detail, full details of the tensor and polarization equations can be found in [6, 8, 9]. These equations we now implement numerically, computing

the power spectrum to an accuracy of about 1%.

Our implementation of the covariant equations is based on the publicly available CMBFAST<sup>1</sup> code [13], itself developed from the COSMICS package<sup>2</sup>. The CMBFAST code uses the line of sight method to achieve high efficiency without compromising accuracy. Our implementation of closed models however differs significantly from CMBFAST version 3.2 described in [76] since our closed code was developed independently. The code described here is up to eight times faster than CMBFAST and has been checked for accuracy over a wide range of model parameters. We refer to this code as *CAMB*, short for “Code for Anisotropies in the Microwave Background”. It was written in Fortran 90 and is now publicly available<sup>3</sup>. It has been used to place good constraints on the curvature of the universe from the BOOMERANG and MAXIMA data [15–18] and by other authors.

Previous results (prior to CMBFAST 3 and *CAMB*) for anisotropies in closed models did not take account of polarization and were obtained by a slow integration of the full Boltzmann hierarchy (see [77]). Using *CAMB* we can compute new results for the scalar and tensor polarization power spectra in closed models. The intensity power spectra for closed models also take account of the polarization and therefore differ from previous results that neglected polarization.

We now describe the key features of the *CAMB* implementation that differ from CMBFAST. A brief description of the code also appears in [78].

## 7.2 Modifying CMBFAST

CMBFAST is written in Fortran 77, an obsolete language that makes writing well-structured, easily-extendible code almost impossible. We devoted a large amount of time to converting the code to Fortran 90 (a slightly improved programming language—though still not object-oriented and with many limitations), removing redundancy and imposing some structure. Once this was completed it was relatively straightforward to implement the covariant equations simply by replacing the synchronous gauge routines of CMBFAST with routines using the covariant variables.

At this point we had an implementation for flat and open models. In order to support closed models we needed to adapt the open code. Modifications to use discrete wavenumbers are straightforward, but computing the ultra-spherical Bessel functions efficiently is rather more complicated. We first adapted the open code using the method described in the next section. We then thoroughly tested the new open code for consistency with the original open code, and only then did we make the further minor modifications needed to support closed models. This ensured that the bulk of our code was thoroughly tested before supporting closed models for which there was no independent check. We could then have a high confidence in our new results for closed models, with the further check that they should agree with the flat results in the zero-curvature limit. All results were then checked for robustness against increases in accuracy parameters (as described later) and accuracy

---

<sup>1</sup><http://www.sns.ias.edu/~matiasz/CMBFAST/cmbfast.html>

<sup>2</sup><http://arcturus.mit.edu/cosmics>

<sup>3</sup><http://www.mrao.cam.ac.uk/~aml1005/cmb>

problems inherited from CMBFAST 2.4.1 were removed. Ultimately the results of the new CMBFAST 3.2 broadly agreed with our code in all models.

### 7.3 Computing the line of sight integral

By integrating the line-of-sight integral (6.176) by parts the integrand becomes a product of a Bessel function with some  $l$ -independent source term. Once the source terms have been computed the line of sight integral is therefore a highly efficient way to compute the anisotropies since the only function that has to be computed for each multipole is a Bessel function. The equations for the multipole hierarchies and line of sight integrals, including tensors and polarization, are collected in Appendices A and B. As can be seen the polarization and tensor multipole follow a similar pattern to the scalar multipoles that we considered in detail in the previous chapter, and the line of sight method can be applied in a similar way. For further details of the tensor and polarization equations see [6, 9].

For closed models we assume that the development angle  $\chi$  satisfies  $\chi < \pi$ , where  $\chi = \eta_0/r$  and  $\eta_0$  is the conformal age of the universe. This permits calculations with models up to  $\Omega_{\text{tot}} \approx 1.7$  for a matter fraction  $\Omega_{\text{m}} \approx 0.4$ . For the case  $\chi \approx \pi$ , all lines of sight converge to the antipodal point close to last scattering, and the power spectrum becomes featureless [77].

#### 7.3.1 Computing the sources

The sources in the line of sight integral depend only on the lowest few multipoles. They can therefore be calculated to good accuracy by propagating a truncated Boltzmann hierarchy. We propagate the hierarchy and background equations using an adaptive Runge-Kutta algorithm as in CMBFAST. To truncate the hierarchy with small errors we use the instantaneous recombination approximation, equation (6.179), to make the approximation  $I_l \propto \Phi_l'$ . Using the relation

$$\frac{d\Phi_l'}{dx} = kr\Phi_{l-1}' - (l+1)\cot_K x\Phi_l' \quad (7.1)$$

for  $l > 0$  we can then determine  $I_l'$  approximately in terms of  $I_l$  and  $I_{l-1}$ . Similar truncation schemes can be used for the neutrino, polarization and tensor hierarchies. At early times the electron density is large and we can use the tight coupling approximation to propagate the lowest photon multipoles—the higher multipoles are strongly suppressed by the scattering. We include the effects of neutrino perturbations when computing the tensor power spectra in order to obtain accurate results for the polarization (CMBFAST neglects the neutrinos as they have a small effect on the temperature power spectrum at low multipoles).

We choose to propagate the equations in the CDM frame, where  $u$  is the velocity of the CDM so that the acceleration  $A = 0$ . We allow for adiabatic, CDM and baryon isocurvature initial conditions as in CMBFAST (modifying to allow for the neutrino isocurvature modes would be straightforward). The equations are propagated forward in time for each wave number starting when all relevant scales are well outside the horizon in the radiation dominated era. Sources are calculated and stored at various time steps up to the present day. Smaller time steps are used over the period of recombination as this is where the largest

contributions come from. The sources vary smoothly and can be interpolated as needed to evaluate them at any given time.

In open and flat models the wavenumbers for which the sources are computed are space logarithmically for low  $k$ , and then in two blocks of increasing linear spacing for higher  $k$ . In closed models the wavenumbers are discretized and the equivalent open/flat values are rounded to the nearest allowed wavenumber. The sources can be interpolated later to evaluate them for each wavenumber required when computing the line of sight integral.

For testing purposes and convenience of other users we include an option to use the synchronous gauge equations [14] originally used by CMBFAST. These are just linear combinations of our CDM frame covariant equations. Given the objective of computing CMB anisotropies the implementation of the covariant equations was strictly redundant. The main reason for the CAMB code being useful is the way that it handles closed and open models, which comes down to how it handles the hyperspherical Bessel functions. This is where the main differences to CMBFAST lie.

### 7.3.2 Evaluating the Bessel functions

The value of the Bessel function at a point can be calculated accurately using the recursion relation (from (6.171) and (6.172))

$$(2l + 1) \cot_K x \Phi_l^\nu = kr (\beta_{l+1} \Phi_{l+1}^\nu + \Phi_{l-1}^\nu) \quad (7.2)$$

for  $l > 0$ . The relation is iterated upwards using analytic forms for the first two functions, or downwards from a guessed (small) value at high  $l$ . For high multipoles a large number of iterations are needed and the evaluation is slow. Two sample Bessel functions<sup>4</sup> are plotted in Figure 7.1.

The differential equation satisfied by the Bessel functions, (6.168) can be written in terms of  $u_l(x) = \sin_K x \Phi_l^\nu(x)$  as

$$\frac{d^2 u_l^\nu}{dx^2} = \left( \frac{l(l+1)}{\sin_K^2 x} - \nu^2 \right) u_l^\nu. \quad (7.3)$$

The hyperspherical Bessel functions can therefore be computed by integrating this equation from initial values for the function and its derivative. Since we are computing an integral over  $x$ , and therefore need the value of the Bessel function for many closely-separated points, this is an efficient way to proceed once the starting values are known.

Kosowsky has developed a WKB approximation to the above differential equation [79]. This gives very accurate results for large  $l$  except in the region  $x \approx \pi/2$  in closed models. Using the WKB approximation when it is accurate, or recursive evaluation otherwise, we can quickly evaluate the Bessel function and its derivative at a starting point. We then integrate the differential equation from that point using the standard fourth-order Runge-Kutta algorithm.

There is however a potential problem. The functions that we want are well behaved everywhere, however there are also solutions which are divergent at  $\sin_K x = 0$ . When

---

<sup>4</sup>For compatibility the code computes Bessel functions with conventional normalization  $\Phi_l^\nu \prod_{n=1}^l \beta_n^{1/2}$ , which are also what is plotted.

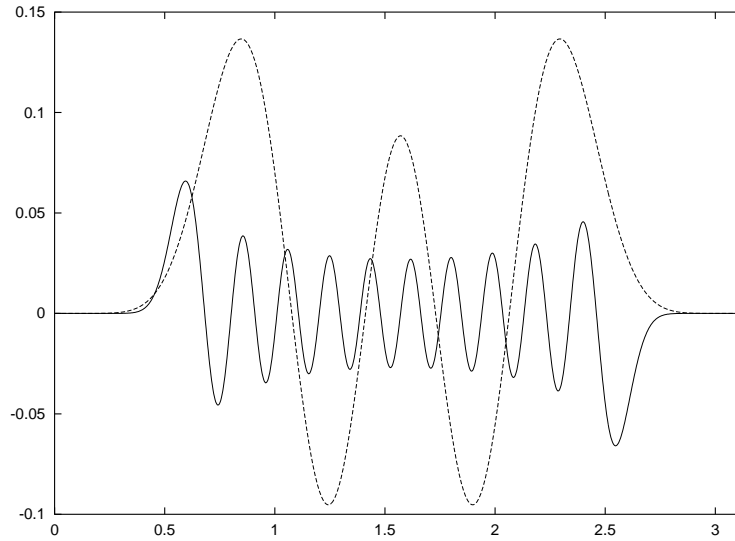


Figure 7.1: Two closed hyperspherical Bessel functions with  $l = 8$ ,  $\nu = 13$  (dashed line) and  $l = 20$ ,  $\nu = 40$  (solid line). The closed functions have  $\nu - l$  extrema, have even or odd symmetry about  $\pi/2$  depending on whether  $l$  is odd or even, and have two regions at either end where the functions become exponentially small.

implementing numerically it is important to avoid contamination with these solutions. For  $\sin_K x$  smaller than the turning point value

$$\sin_K x_t = \frac{\sqrt{l(l+1)}}{\nu} \quad (7.4)$$

the equation becomes dissipative and  $u_l'$  gets exponentially small. It is in this region that we must be careful to avoid contamination with the irregular solution that will come to dominate any numerical integration of the differential equation. In open models this only occurs once and it is possible to integrate up from some point where the Bessel function is very small. However in closed models there can be two dissipative regions if the development angle exceeds  $\pi/2$ . Integrating up from the first region will then eventually lead to a problem in the second region. Figure 7.2 shows what happens if we numerically integrate up and down from initial values at the turning point. Exactly similar behaviour is found integrating down and up from the turning point in the  $\chi > \pi/2$  region.

It is clear from the figure that the problem only becomes significant once the numerical value of the Bessel function becomes small. However since the contribution to the line-of-sight integral becomes small when  $\Phi_l'$  is small we can simply cut-off the integration at a suitable point. We start the integration at around the turning point and integrate up and down from there, cutting-off when the irregular solution contamination causes the derivative of  $u_l'$  to change sign in the dissipative regions. Figure 7.3 shows the exact solution compared with the values obtained by numerical integration down to the cut-off. As we can see the

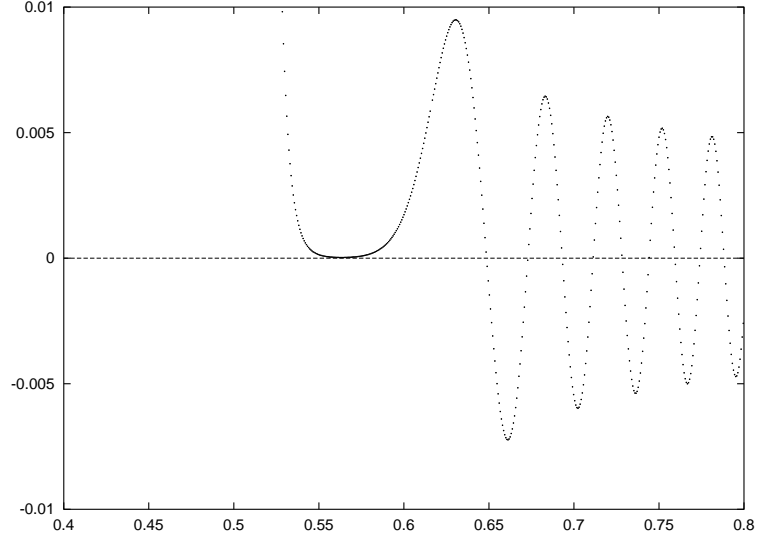


Figure 7.2: The values of  $\Phi_l'(\chi)$  obtained by numerically integrating the differential equation up and down from the inflection point for an  $l = 224$ ,  $\nu = 388$  closed hyperspherical Bessel function.

two only differ significantly once the Bessel function has fallen to less than 1% of its peak value, and numerical trials confirm that using the simple cut-off gives errors consistent with the target 1% accuracy in the power spectrum. It would be a simple matter to devise more accurate schemes, however we opt for the simplest solution that leads to accurate results.

Using the recursion relation and WKB approximation to quickly evaluate the Bessel functions at a point allows us to skip ranges of the integration which are insignificant and re-start again when it becomes important again. For example in many cases the only significant contributions arise from the sources around recombination and the entire integral down the line of sight to the last scattering surface can be safely neglected. Skipping these negligible contributions can give a significant reduction in computation time.

In CMBFAST 3 the Bessel functions are evaluated by integrating the differential equation up from a pre-computed starting point. The problem with irregular-solution contamination is avoided by using the symmetries of the Bessel functions to find values in the second region from values generated by upwards integration in the first region. The disadvantage of this approach is that it prevents a simple scheme for skipping the regions that give negligible contributions to the integral. The straightforward integration with cut-off in closed models used by CAMB makes the code significantly faster, and well as being much simpler, making the code more transparent and less susceptible to coding errors.



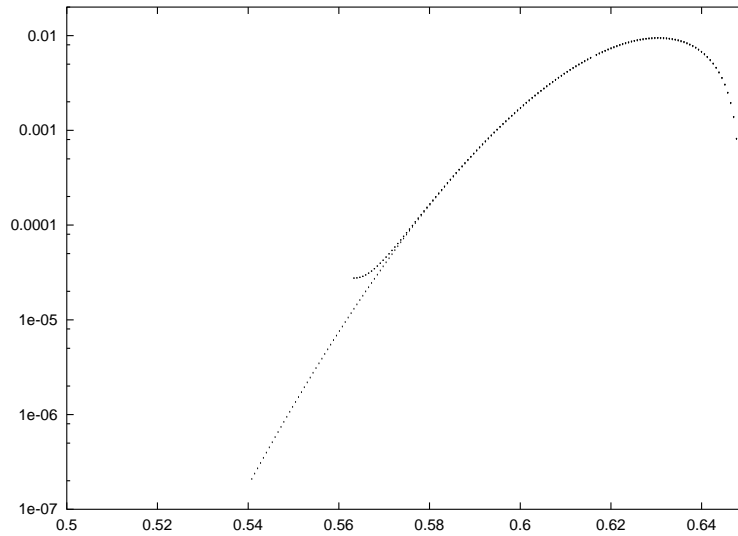


Figure 7.3: The exact regular solution (small dots) and the numerically integrated solution with cut-off (large dots) for an  $l = 224$ ,  $\nu = 388$  closed hyperspherical Bessel function. A log scale has been used to make the error visible — the integrated solution is very accurate until the function is less than 1% of the peak height.

## 7.4 The power spectrum: integrating over wavenumber

The wavenumbers sampled in order to perform the numerical integration are logarithmically spaced for low  $k$  and then linearly spaced for higher  $k$  as for the  $k$ -sampling when computing the sources. For closed models the integral is replaced by a sum and we can use every allowed discrete wavenumber. However for nearly flat closed models the number of discrete wavenumbers we would have to sum is large ( $k \sim \nu/r$  where  $r$  is large and  $\nu$  is integer). When the model is nearly flat we therefore sample the wavenumbers as for open and flat models (but rounding to the nearest allowed wavenumber, and decreasing the step sizes to allow for aliasing effects).

## 7.5 Massive neutrinos

Massive neutrinos are handled by propagating the energy dependent equations whilst the neutrinos are relativistic and then switching to a truncated energy-integrated hierarchy. We propagate the momentum-dependent scalar multipole equations (6.208) for 15 co-moving momenta to give accurate numerical integration over the Fermi-Dirac distribution function (as in CMBFAST). We truncate the hierarchy as in the massless case, though at a lower multipole since the mass suppresses the higher multipoles. The terms in the neutrino stress-energy tensor are computed as required by integrating numerically over momentum.

When the momentum has redshifted so that the neutrinos are no longer highly relativistic-

tic we switch to propagating the truncated energy-integrated hierarchy. Keeping terms up to  $l + 2i = 3$  we have the six equations

$$\chi' + 3H(\chi + \chi^p) + 3h'(\rho + p) + kJ_1 = 0 \quad (7.5)$$

$$\chi^{p'} + 5H\chi^p + 5h'p - \frac{1}{3}kJ_1^{(1)} = 0 \quad (7.6)$$

$$J_1' + 4HJ_1 + k[\frac{2}{3}\beta_2J_2 - \chi^p] = 0 \quad (7.7)$$

$$J_2' + 5HJ_2 + k[\frac{3}{5}\beta_3J_3 - \frac{2}{5}J_1^{(1)}] - 2k\sigma p = 0 \quad (7.8)$$

$$J_3' + 6HJ_3 = 0 \quad (7.9)$$

$$J_1^{(1)'} + 6HJ_1^{(1)} = 0. \quad (7.10)$$

The last two have the analytic solutions

$$J_3 \propto S^{-6} \quad J_1^{(1)} \propto S^{-6} \quad (7.11)$$

so to this order we need to propagate only four perturbation equations. The starting values can be computed by numerically integrating over momentum at the switch-over point. Since the higher order variables are most important on small scales the switch over point needs to be rather later for high  $k$  modes. Switching to the energy-integrated equations nearly halves the computation time compared with using just the momentum-dependent equations as in previous accurate codes [12, 13].

Recent evidence from oscillation of atmospheric neutrinos provides evidence for a small mass difference between neutrinos<sup>5</sup>. This either implies that the neutrinos are all very light, or that the masses are nearly degenerate. In the computation we allow for  $N_\nu$  species of massive neutrino of degenerate mass (as in CMBFAST). This should be a good approximation for neutrinos masses that have any cosmological significance as very light non-degenerate masses would contribute only about  $\Omega_\nu \sim 0.001$  [80]. Ultimately data from the CMB may be able to provide good bounds on the actual magnitude of the neutrino masses rather than just mass differences deduced from other experiments.

## 7.6 Numerical stability and accuracy

We have a variety of non-independent variables to choose from when we propagate the equations. It is important to maintain numerical stability in the differential equations for the dependent variables. This is especially true for scalar perturbations with isocurvature initial conditions, where a poor choice of dependent variables can lead to large violations of the constraint equations. Numerical studies show that propagating  $\eta$ , the gradient of the 3-Ricci curvature, and  $\sigma$ , the shear, give good numerical stability for all initial conditions. If we instead propagate  $\mathcal{Z}$ , the gradient of the expansion, the equations are unstable for isocurvature initial conditions.

<sup>5</sup><http://www.hep.anl.gov/ndk/hypertext/nuindustry.html>

We introduce various global parameters that control the accuracy of the computations — controlling the time steps,  $k$ -sampling, etc. It is thus possible to compute power spectra with different settings and thereby assess the accuracy of the computation. We computed power spectra for a grid of different models and adjusted the accuracy parameters so that the results agreed to better than 1% with results generated with much higher accuracy settings.

To give an accurate computation of the ionization history of the universe we allow use of RECFAST, described in [81]. This corrects errors at about the 2% level.

CMBFAST 3.2 gives inaccurate results for the tensor quadrupole in flat models [82]. We fix this by computing the Bessel functions directly where errors in CMBFAST’s Bessel function interpolation are important.

The power spectrum computed will of course depend on the initial power spectrum. We use essentially the same driver routine as CMBFAST to allow for the most frequently used spectra and to be consistent with the parameters expected by CMBFAST. However as these are only accurate for a small number of inflationary models [82] we allow for a custom initial power spectrum so the code can be used for more general models.

## 7.7 Interpolating the power spectrum

In closed models a given linear scale at last scattering subtends a larger angle on the sky today than in open or flat models. Conversely scales in open models subtend smaller angles. This geometric effect shifts power in the CMB spectra to smaller  $l$  in closed models and larger  $l$  in open models. This effect is generally much larger than shifts due to a change in sound horizon at last scattering that can result from varying the other parameters. Since CMBFAST only computes the power spectra at a few values of  $l$  and interpolates between them we need to adjust the  $l$ -sampling according to the curvature to maintain accurate interpolation. We calculate the shift in the power spectrum due to the curvature for each model and adjust the multipoles at which the  $C_l$  are computed dynamically. This contrasts with the cumbersome static approach of CMBFAST 3.2.

## 7.8 Results

We have verified our calculations against results obtained with CMBFAST version 3.2 with good agreement well into the damping tail.

In Figure 7.4 we plot the intensity and polarization power spectra in  $\Lambda$ CDM models assuming no reionization. The power spectra for the polarization and tensor modes are defined analogously to (6.197), for details see [9]. One model is closed ( $\Omega_{\text{tot}} = 1.2$ ,  $\Omega_{\Lambda} = 0.8$ ), while the other is flat ( $\Omega_{\text{tot}} = 1$ ,  $\Omega_{\Lambda} = 0.6$ ). In both cases we take the matter fraction  $\Omega_{\text{m}} = 0.4$ , baryon fraction  $\Omega_{\text{b}} = 0.045$ , and Hubble’s constant  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . For the scalar modes, we assume adiabatic initial conditions with a scale-invariant initial power spectrum.

Since the cosmological constant and curvature are unimportant at early times, and the physical matter densities  $\Omega_{\text{m}} H_0^2$  and  $\Omega_{\text{b}} H_0^2$  are the same, the two models evolve the

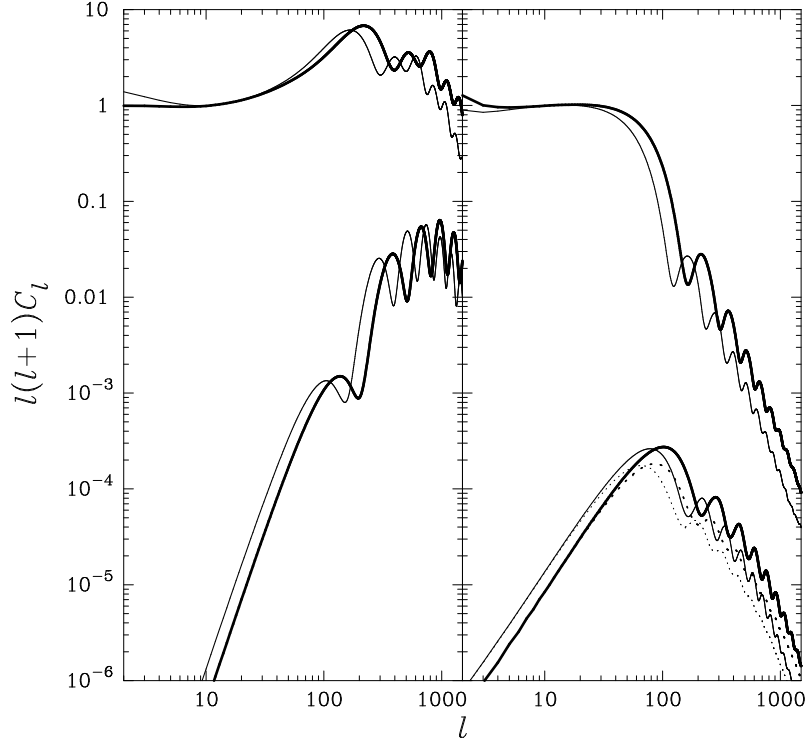


Figure 7.4: Scalar (left) and tensor (right) intensity and polarization power spectra in a closed CDM model ( $\Omega_{\text{tot}} = 1.2$ ,  $\Omega_{\Lambda} = 0.8$ ; thin lines), and a flat model ( $\Omega_{\text{tot}} = 1$ ,  $\Omega_{\Lambda} = 0.6$ ; thick lines). In both cases  $\Omega_{\text{m}} = 0.4$ ,  $\Omega_{\text{b}} = 0.045$ , and  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The upper solid lines are the intensity, the lower ones the electric component of the polarization  $C_l^{\mathcal{E}\mathcal{E}}$ , and the dashed lines the magnetic component  $C_l^{\mathcal{B}\mathcal{B}}$ . The scalar and tensor intensities are normalized to unity at  $l = 10$ . There is no scalar magnetic polarization as it has no source terms for a scalar mode.

same way until last scattering. There are two main differences caused by difference in curvature. The shift in the spectrum comes from the different angular diameter distances to the last scattering surface caused by the geometry—in the closed model the geodesics converge so features on the last scattering surface appear at larger angular scales on the sky. The differences at low  $l$  arises from the integrated Sachs-Wolfe effect—the variation of the potentials after last scattering given by the time integral in (6.177).

## 7.9 Conclusion

We have implemented the covariant cosmological perturbation equations numerically, producing accurate and fast code. The code has been used in the extraction of cosmological parameters from observational data given certain assumptions about the initial power spectrum. Ultimately it should be of use for reconstructing the initial power spectrum and constraining the inflationary potential from forthcoming high-quality CMB data<sup>6</sup>, especially in conjunction with more accurate determinations of cosmological parameters from other independent experiments.

---

<sup>6</sup>For example MAP (<http://map.gsfc.nasa.gov>) and Planck (<http://astro.estec.esa.nl/SA-general/Projects/Planck>)

### Appendix A: The Scalar Equations

The full intensity and electric polarization<sup>1</sup> hierarchies are:

$$I'_l + k \left\{ \frac{l+1}{2l+1} \beta_{l+1} I_{l+1} - \frac{l}{2l+1} I_{l-1} \right\} =$$

$$-4h'\delta_l^0 - 4kA\delta_l^1 + \frac{8}{15}\sigma\delta_l^2 - S n_e \sigma_T (I_l - \delta_l^0 I_0 - \frac{4}{3}\delta_l^1 v - \frac{2}{15}\zeta\delta_l^2)$$

$$\mathcal{E}'_l + k \left\{ \frac{(l+3)(l+1)}{(2l+1)(l+1)} \beta_{l+1} \mathcal{E}_{l+1} - \frac{l}{2l+1} \mathcal{E}_{l-1} \right\} = -S n_e \sigma_T (\mathcal{E}_l - \frac{2}{15}\delta_{2l}\zeta).$$

The line of sight integral formulae are:

$$I_l|_{\eta_0} = 4 \int^{\eta_0} d\eta e^{-\tau} \left\{ (k\sigma + \frac{1}{4}S n_e \sigma_T \zeta) \left[ \frac{1}{3}\Phi'_l(\chi) + \frac{1}{k^2 r^2} \frac{d^2}{d\chi^2} \Phi'_l(\chi) \right] \right.$$

$$\left. - (kA - S n_e \sigma_T v) \frac{1}{kr} \frac{d}{d\chi} \Phi'_l(\chi) + \left[ \frac{1}{4}S n_e \sigma_T I_0 - h' \right] \Phi'_l(\chi) \right\}$$

$$\mathcal{E}_l|_{\eta_0} = \frac{l(l-1)}{k^2 r^2} \int^{\eta_0} d\eta S n_e \sigma_T e^{-\tau} \zeta \frac{\Phi'_l(\chi)}{\sin_K^2 \chi}$$

### Appendix B: The Tensor Equations

For tensors define

$$\beta_l \equiv 1 - (l^2 - 3) \frac{K}{k^2} \quad \nu^2 \equiv \frac{k^2 + 3K}{|K|}.$$

The intensity, electric and magnetic polarization<sup>1</sup> hierarchies are

$$I'_l + k \left\{ \frac{(l+3)(l+1)}{(2l+1)(l+1)} \beta_{l+1} I_{l+1} - \frac{l}{2l+1} I_{l-1} \right\} - \frac{8}{15} k \sigma \delta_{2l} = -S n_e \sigma_T (I_l - \frac{2}{15}\zeta\delta_{2l})$$

$$\mathcal{E}'_l + k \left\{ \frac{(l+3)^2(l+1)^2}{(2l+1)(l+1)^2} \beta_{l+1} \mathcal{E}_{l+1} - \frac{l}{2l+1} \mathcal{E}_{l-1} \right\} - \frac{4k}{l(l+1)} \beta_0^{1/2} \bar{\mathcal{B}}_l = -S n_e \sigma_T (\mathcal{E}_l - \frac{2}{15}\zeta\delta_{2l})$$

$$\mathcal{B}'_l + k \left\{ \frac{(l+3)^2(l+1)^2}{(2l+1)(l+1)^2} \beta_{l+1} \mathcal{B}_{l+1} - \frac{l}{2l+1} \mathcal{B}_{l-1} \right\} + \frac{4k}{l(l+1)} \beta_0^{1/2} \bar{\mathcal{E}}_l = -S n_e \sigma_T (\mathcal{B}_l - \frac{2}{15}\zeta\delta_{2l})$$

and the line of sight integral formulae are

$$I_l|_{\eta_0} = \frac{l(l-1)\beta_1^{-1/2}}{k^2 r^2} \int^{\eta_0} d\eta (k\sigma + \frac{1}{4}S n_e \sigma_T \zeta) \frac{\Phi'_l(\chi)}{\sin_K^2 \chi}$$

$$\mathcal{E}_l|_{\eta_0} = \frac{l(l-1)\beta_1^{-1/2}}{k^2 r^2 (l+1)(l+2)} \int^{\eta_0} d\eta e^{-\tau} S n_e \sigma_T \zeta \left[ \frac{d^2}{d\chi^2} + 4 \cot_K \frac{d}{d\chi} + 2(1 + \cot_K^2) - k^2 r^2 \right] \Phi'_l(\chi)$$

$$\mathcal{B}_l|_{\eta_0} = \frac{-2\nu l(l-1)\beta_1^{-1/2}}{k^2 r^2 (l+1)(l+2)} \int^{\eta_0} d\eta e^{-\tau} S n_e \sigma_T \bar{\zeta} \left[ \frac{d}{d\chi} + 2 \cot_K \right] \Phi'_l(\chi).$$

(bars here denote magnetic parity variables). For derivations, further references, and a discussion of the mode expansion and power spectra see [8, 9].

<sup>1</sup>The definitions of the polarization harmonic expansion variables include an energy density factor and minus sign, so for example  $\zeta = 3/4I_2 + 2/9\mathcal{E}_2$ .

## Chapter 8

# Conclusions

In Chapter 2 we showed how GA could be used as a powerful tool for co-ordinate free manipulations of geometric objects. Geometric calculus allowed for an extension of traditional tensor calculus and we developed a notation that allows our results to be related easily to those using traditional methods. We then applied GA to a variety of topics, demonstrating the advantages of the GA approach in several areas.

In Chapter 3 we showed how Dirac theory can be given a physically transparent GA formulation. We showed how the theory could be applied to single particle scattering, handling spin-dependence in a natural way. Spin orientations were handled directly and we had no need to perform spin-sums or introduce the abstract gamma-matrices used in other approaches. We attempted to study more general scattering calculations involving more than one particle with some success. However multi-particle constructions in GA seem rather forced, and this remains an area where traditional methods still have some advantage. It would be interesting to see whether quantum field theories could be studied in GA maintaining the advantages of the single particle formalism. However at the multi-particle level things are rather more complicated and it is not clear whether the GA approach would offer any advantages over traditional methods.

In the remaining chapters we considered various topics in gravitational physics. In Chapter 4 we showed how gauge theory gravity could be formulated in GA, and discussed an extension to a scale invariant theory. In most cases this reduces to an extension of GR with a massive vector field and seems to have limited interest, though suggesting yet another possible dark matter candidate. It remains to be seen whether local scale invariance is a useful idea or not, and future work could consider the quantum implications and interactions with the standard model fields.

The gauge theory formulation of gravity allows for significant freedom in the choice of action, even with the addition of scale invariance. In Chapter 5 we showed that instanton configurations of the rotation gauge field gave rise to topological terms in the action. We constructed the independent quadratic terms for the action and calculated the modified field equations for some of the possible terms. Although we found the field equations we did not perform any further analysis, largely because of the complexity and number of the equations. It would be interesting to look at the effect of the additional terms on the gravitational force

law, and in particular at the effects in cosmology and black hole formation. If the current observational disagreements with the CDM cosmological model are confirmed, and other forms of dark matter do not form a successful substitute, it may become a pressing matter to consider different gravitational theories. The issue of quantization of gauge theory gravity should also be addressed.

For the moment General Relativity still appears to be an accurate theory, and so we expect any extended gravitational theory to tend to GR for most cases of observational interest. In the remaining chapters we assumed that GR was a sufficiently accurate theory and studied the propagation of covariant perturbations in cosmology. By construction the covariant perturbation variables are observables, and the covariant analysis in Chapter 6 gave a physically transparent derivation of the propagation equations. We extended previous results to include massive neutrinos and studied the Boltzmann equation via a two dimensional hierarchy of energy-integrated multipole equations. We also discussed the propagation of covariant perturbations during inflation, and related the equations to the well known gauge-invariant equations.

At the moment there is great interest in the measurement of CMB anisotropies, and it is vital to have accurate theoretical predictions to compare with observations. In Chapter 7 we implemented the mode expanded covariant perturbation numerically to compute predictions for the CMB power spectrum given a set of cosmological parameters and an initial power spectrum. The code supports closed models and computes accurate results with high efficiency. We were able to compute new predictions for CMB anisotropies in closed models taking full account of polarization. The code was made publicly available and has already proved useful in the extraction of cosmological parameters from observational data and in constraining the curvature of the universe.

Forthcoming high-precision data should determine many parameters with unprecedented accuracy and enable a reconstruction of the initial power spectrum. Ultimately we would like to learn about the period of inflation, and to this end it would be useful to construct code for computing the initial power spectrum from an inflationary model. We have already derived the propagation equations for the inflaton perturbations and shown how they relate to the initial power spectrum. Implementing the equations numerically could give accurate predictions for the initial power spectrum without having to use the slow-roll or other approximations. The new high-precision CMB data should then be able to pin down the inflationary theory to some limited accuracy. It may also be possible to study the generation of perturbations from a primordial instanton to constrain the ultimate origin of the universe. Much work remains to be done however before the covariant approach can be used to study quantum effects, and there remains the question of which gravitational theory is the correct one to quantize.

There are a many effects that we have not considered that can be significant to CMB anisotropy measurements. In particular we have ignored gravitational lensing and the non-linear effects than can become important at high angular resolutions. The code we have developed could be extended to take these into account, and a covariant analysis would provide a useful check on previous results. There is also the possibility of quintessence, warm and interacting dark matter, and a wide variety of other variations on the standard cosmological model. To test these models accurately against observations we shall need



robust theoretical models and accurate numerical codes to take the new effects into account. The covariant approach to cosmological perturbations can readily be extended to study non-linear effects, and should prove useful in high-resolution studies of CMB and in the evolution of non-linear structures.

Overall we have shown that various topics can be handled in a physical co-ordinate independent way using Geometric Algebra and covariant methods. Whilst we do not propose a dogmatic adherence to a particular mathematical approach these methods appear to be a good way to tackle a variety of physical problems. Cosmology is at a turning point — within the next decade we should have accurate data that could tell us the main cosmological parameters with great accuracy. There remains much work to be done to ensure that the theoretical predictions are robust, and to enable the observations to constrain the inflationary model and matter content of the universe. Covariant methods should prove a useful tool for tackling this work.



# Bibliography

- [1] D. Hestenes and G. Sobczyk. *Clifford Algebra to Geometric Calculus*. Reidel, Dordrecht, 1984.
- [2] D. Hestenes. *Space-Time Algebra*. Gordon and Breach, New York, 1966.
- [3] D. Hestenes. Geometry of the Dirac theory. In J. Keller, editor, *The Mathematics of Physical Spacetime*, page 67. UNAM, Mexico, 1982.
- [4] A. Lewis. Elastic potential scattering in the spacetime algebra. unpublished, 1997.
- [5] A.N. Lasenby, C.J.L. Doran, and S.F. Gull. Gravity, gauge theories and geometric algebra. *Phil. Trans. R. Soc. Lond. A*, 356:487–582, 1998.
- [6] A. Challinor. *Geometric Algebra and its Applications in physics and cosmology*. PhD thesis, Cambridge University, 1998.
- [7] A. Challinor and A. Lasenby. Cosmic microwave background anisotropies in the cold dark matter model: A covariant and gauge-invariant approach. *astro-ph/9804301*. *Ap. J.*, 513(1):1–22, 1999.
- [8] A. Challinor. Microwave background anisotropies from gravitational waves: the 1+3 covariant approach. *astro-ph/9906474*. *Class.Quant.Grav.*, 17:871–889, 2000.
- [9] A. Challinor. Microwave background polarization in cosmological models. *astro-ph/9911481*, 1999.
- [10] G.F.R. Ellis, D.R. Matravers, and R. Treciokas. Anisotropic solutions of the Einstein-Boltzmann equations: I. General formalism. *Ann. Phys.*, 150:455, 1983.
- [11] V.F. Mukhanov, H.A. Feldman, and R.H. Brandenberger. Theory of cosmological perturbations. *Phys. Rep.*, 215:203–333, 1992.
- [12] C.P. Ma and E. Bertschinger. Cosmological perturbation theory in the synchronous and conformal Newtonian gauges. *astro-ph/9506072*. *Ap. J.*, 455:7, 1995.
- [13] U. Seljak and M. Zaldarriaga. A line of sight integration approach to Cosmic Microwave Background anisotropies. *astro-ph/9603033*. *Ap. J.*, 469(2):437–444, 1996.

- [14] M. White W. Hu, U. Seljak and M. Zaldarriaga. Complete treatment of CMB anisotropies in a FRW universe. *Phys. Rev. D*, 57(6):3290–3301, 1998.
- [15] A. Melchiorri et al. A measurement of Omega from the North American test flight of BOOMERANG. *astro-ph/9911445. Ap. J*, 536:L63–L66, 2000.
- [16] P. de Bernardis et. al. A flat universe from high-resolution maps of the cosmic microwave background radiation. *astro-ph/0004404. Nature*, 404:955–959, 2000.
- [17] A.E. Lange et. al. First estimates of cosmological parameters from BOOMERANG. *astro-ph/0005004*, 2000.
- [18] A. Balbi et. al. Constraints on cosmological parameters from MAXIMA-1. *astro-ph/0005124*, 2000.
- [19] S.F. Gull, A.N. Lasenby, and C.J.L. Doran. Imaginary numbers are not real — the geometric algebra of spacetime. *Found. Phys.*, 23(9):1175, 1993.
- [20] C.J.L. Doran, A.N. Lasenby, and S.F. Gull. States and operators in the spacetime algebra. *Found. Phys.*, 23(9):1239, 1993.
- [21] A.N. Lasenby, C.J.L. Doran, and S.F. Gull. A multivector derivative approach to Lagrangian field theory. *Found. Phys.*, 23(10):1295, 1993.
- [22] C.J.L. Doran, A.N. Lasenby, S.F. Gull, S.S. Somaroo, and A.D. Challinor. Spacetime algebra and electron physics. *Adv. Imag. & Elect. Phys.*, 95:271, 1996.
- [23] R.H. Dalitz. On higher born approximation in potential scattering. *Proc. Roy. Soc*, 206:509, 1951.
- [24] C. Itzykson and J-B. Zuber. *Quantum Field Theory*. McGraw-Hill, New York, 1980.
- [25] N.F. Mott. Scattering of fast electrons by atomic nuclei. *Proc. Roy. Soc*, A124:425, 1929.
- [26] N.F. Mott. Polarization of electrons by double scattering. *Proc. Roy. Soc*, A315:429, 1932.
- [27] J.D. Bjorken and S.D. Drell. *Relativistic Quantum Mechanics, vol 1*. McGraw-Hill, New York, 1964.
- [28] R.P. Feynman, F.B. Morningo, and W.G. Wagner. *Feynman Lectures on Gravitation*. Addison-Wesley, Reading MA, 1995.
- [29] D. Sciama. The physical structure of general relativity. *Rev. Mod. Phys.*, 36:463 and 1103, 1964.
- [30] E. Cartan. Sur un généralisation de la notion de courbure de Riemann et les espaces à torsion. *C. R. Acad. Sci. (Paris)*, 174:593, 1922.

- [31] T.W.B. Kibble. Lorentz invariance and the gravitational field. *J. Math. Phys.*, 2(3):212, 1961.
- [32] T. Eguchi, P.B. Gilkey, and A.J. Hanson. Gravitation, gauge theories and differential geometry. *Phys. Rep.*, 66(6):213, 1980.
- [33] D. Ivanenko and G. Sardanashvily. The gauge treatment of gravity. *Phys. Rep.*, 94(1):1, 1983.
- [34] F.W. Hehl and B.K. Datta. Nonlinear spinor equation and asymmetric connection in general relativity. *J. Math. Phys.*, 12(7):1334, 1971.
- [35] M. Gockeler and T. Schucker. *Differential Geometry, Gauge Theories, and Gravity*. Cambridge University Press, 1987.
- [36] P.A.M. Dirac. Long range forces and broken symmetries. *Proc. Roy. Soc. (London)*, A333:403, 1973.
- [37] M. Israelit. Massive electrodynamics and magnetic monopoles. *GR and Grav.*, 29(11):1411, 1997.
- [38] M. Israelit and N. Rosen. Weyl-Dirac geometry and dark matter. *Found. Phys.*, 22(4):555, 1992.
- [39] N. Rosen. Weyl's geometry and physics. *Found. Phys.*, 12(3):213, 1982.
- [40] D. Gregorash and G. Papini. Weyl-Dirac theory with torsion. *Nuovo. Cim.*, 55B:37, 1980.
- [41] M. Pawłowski and R. Raczka. A Higgs-free model for fundamental interactions. *Found. Phys.*, 24:1305, 1994.
- [42] H. Cheng. Possible existence of Weyl's vector meson. *Phys. Rev. Lett.*, 61(19):2182, 1988.
- [43] W. Drechsler and H. Tann. Broken Weyl-invariance and the origin of mass. *gr-qc/9802044*, 1998.
- [44] J. Solà. The cosmological constant and the fate of the cosmon in Weyl conformal gravity. *Phys. Lett. B*, 228(3):317, 1989.
- [45] F.W. Hehl, J.D. McCrea, E.W. Mielke, and Y. Ne'eman. Metric-affine gauge theory of gravity: field equations, Noether identities, world spinors and breaking of dilaton invariance. *Phys. Rep.*, 258:1–171, 1995.
- [46] F. Hoyle, G. Burbidge, and J. Narlikar. *A Different Approach to Cosmology*. CUP, 2000.
- [47] N. Tsamis and R. Woodward. No new physics in conformal scalar-metric theory. *Ann. Phys.*, 168:257, 1986.

- [48] Y.N. Obukhov and E.J. Vlachynsky. Einstein-Proca model: spherically symmetric solutions. gr-qc/0004081, 2000.
- [49] U. Lindström. The Palatini variational principle and a class of scalar-tensor theories. *Nuovo Cim.*, 35B:130, 1976.
- [50] K.S. Stelle. Classical gravity with higher derivatives. *Gen. Relativ. Gravit.*, 9:353, 1978.
- [51] H.T. Nieh. Gauss-Bonnet and Bianchi identities in Riemann-Cartan type gravitational theories. *J. Math. Phys.*, 21(6):1439, 1980.
- [52] M.A.J. Ashdown, S.S. Somaroo, S.F. Gull, C.J.L. Doran, and A.N. Lasenby. Multilinear representations of rotation groups within geometric algebra. *J. Math. Phys.*, 39(3):1566, 1998.
- [53] S. Weinberg. *The Quantum Theory of Fields, Volume 1: Foundations*. CUP, 1995.
- [54] H.T. Nieh and M.L. Yan. An identity in Riemann-Cartan geometry. *J. Math. Phys.*, 23:373, 1982.
- [55] F. Sommen. Monogenic functions of higher spin. *Z. A. A.*, 15(2):279–282, 1996.
- [56] Yu.N. Obukhov, V.N. Ponomarev, and V.V. Zhytnikov. Quadratic Poincaré gauge theory of gravity: A comparison with General Relativity theory. *Gen. Relativ. Grav.*, 21(11):1107, 1989.
- [57] M.S. Longair. *Galaxy Formation*. Springer, 1998.
- [58] J. Peacock. *Cosmological Physics*. CUP, 1999.
- [59] T. Gebbie and G. Ellis. 1+3 covariant cosmic microwave background anisotropies I: Algebraic relations for mode and multipole representations. *astro-ph/9804316*, 1999.
- [60] T. Gebbie, P.K.S. Dunsby, and G.F.R. Ellis. 1+3 covariant cosmic microwave background anisotropies. II: The almost-Friedmann Lemaitre model. *astro-ph/9904408*, 1999.
- [61] G.F.R. Ellis and H. van Elst. Cosmological models. In M Lachièze-Rey, editor, *Theoretical and Observational Cosmology*, pages 1–116. Kluwer: Dordrecht, 1999.
- [62] R. Maartens, J. Triginer, and D.R. Matravers. Stress effects in structure formation. *astro-ph/9901213*, 1999.
- [63] P. Colín, V. Avila-Reese, and O. Valenzuela. Substructure and halo density profiles in a warm dark matter cosmology. *astro-ph/0004115*, 2000.
- [64] C.J. Hogan and J.J. Dalcanton. New dark matter physics: Clues from halo structure. *astro-ph/0002330*, 2000.

- [65] S. Hannestad. Can dark matter see itself? *astro-ph/0008422*, 2000.
- [66] J.A. Sellwood and A. Kosowsky. Does dark matter exist? *astro-ph/0009074*, 2000.
- [67] R. Bousso and Andrei Linde. Quantum creation of a universe with  $\omega \neq 1$ : Singular and non-singular instantons. *astro-ph/9803068*, *Phys. Rev. D*, 58:083503, 1998.
- [68] M. Bruni, G.F.R. Ellis, and P.K.S. Dunsby. Gauge-invariant perturbations in a scalar field dominated universe. *Class. Quant. Grav.*, 9:921–945, 1992.
- [69] A. Liddle and D. Lyth. The cold dark matter density perturbation. *astro-ph/9303019*, *Phys.Rept.*, 231:1–105, 1993.
- [70] W. Hu and N. Sugiyama. Anisotropies in the microwave background: an analytic approach. *astro-ph/9407093*. *Ap. J.*, 444:489, 1995.
- [71] D. Lyth and A. Woszczyna. Large scale perturbations in the open universe. *astro-ph/9501044*. *Phys. Rev. D*, 52:3338, 1995.
- [72] A. Challinor. Regular initial conditions for scalar perturbations. unpublished, 1999.
- [73] M. Bucher, K. Moodley, and N. Turok. The general primordial cosmic perturbation. *astro-ph/9904231*, 1999.
- [74] M. Bucher, K. Moodley, and N. Turok. Characterising the primordial cosmic perturbations using MAP and PLANCK. *astro-ph/0007360*, 2000.
- [75] S. Dodelson, E. Gates, and A. Stebbins. Cold + hot dark matter and the cosmic microwave background. *astro-ph/9509147*, 1995.
- [76] M. Zaldarriaga and U. Seljak. CMBFAST for spatially closed universes. *astro-ph/9911219*, 1999.
- [77] M. White and D. Scott. Why not consider closed universes? *Ap. J.*, 459(2):415–431, 1996.
- [78] A. Lewis, A. Challinor, and A. Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *astro-ph/9911177*. *Ap. J.*, 538:473, 2000.
- [79] A. Kosowsky. Efficient computation of hyperspherical Bessel functions. *astro-ph/9805173*, 1998.
- [80] J. Primack and M. Gross. Hot dark matter in cosmology. *astro-ph/0007165*, 2000.
- [81] S. Seager, D.D. Sasselov, and D. Scott. A new calculation of the recombination epoch. *astro-ph/9909275*, 1999.
- [82] J. Martin, A. Riazuelo, and D. Schwarz. Slow-roll inflation and CMB anisotropy data. *astro-ph/0006392*, 2000.