Natural observer fields and redshift

ROBERT S MACKAY
COLIN ROURKE

We sketch the construction of space-times with observer fields which have redshift satisfying Hubble's law but no big bang.

83F05; 83C15, 83C40, 83C57

1 Introduction

This paper is part of a program whose aim is to establish a new paradigm for the universe in which there is no big bang. There are three pieces of primary evidence for the big bang: the distribution of light elements, the cosmic microwave background, and redshift. This paper concentrates on redshift. Alternative explanations for the other pieces of primary evidence are given in other papers in the program.

The context for our investigation of redshift is the concept of an *observer field* by which we mean a future-pointing time-like unit vector field. An observer field is *natural* if the integral curves (field lines) are geodesic and the perpendicular 3–plane field is integrable (giving normal space slices). It then follows that the field determines a local coherent notion of time: a time coordinate that is constant on the perpendicular space slices and whose difference between two space slices is the proper time along any field line. This is proved in [6, Section 5].

In a natural observer field, red or blueshift can be measured locally and corresponds precisely to expansion or contraction of space measured in the direction of the null geodesic being considered (this is proved in Section 2 of this paper). Therefore, if one assumes the existence of a global natural observer field, an assumption made implicitly in current conventional cosmology, then redshift leads directly to global expansion and the big bang. But there is no reason to assume any such thing and many good reasons not to do so. It is commonplace observation that the universe is filled with heavy bodies (galaxies) and it is now widely believed that the centres of galaxies harbour supermassive objects (normally called black holes). The neighbourhood of a black hole is not covered by a natural observer field. One does not need to assume that there is a singularity at the centre to prove this. The fact that a natural observer field admits a coherent time contradicts well known behaviour of space-time near an event horizon.

In this paper we shall sketch the construction of universes in which there are many heavy objects and such that, outside a neighbourhood of these objects, space-time admits natural observer fields which are roughly expansive. This means that redshift builds up along null geodesics to fit Hubble's law. However there is no global observer field or coherent time or big bang. The expansive fields are all balanced by dual contractive fields and there is in no sense a global expansion. Indeed, as far as this makes sense, our model is roughly homogeneous in both space and time (space-time changes dramatically near a heavy body, but at similar distances from these bodies space-time is much the same everywhere).

A good analogy of the difference between our model and the conventional one is given by imagining an observer of the surface of the earth on a hill. He sees what appears to be a flat surface bounded by a horizon. His flat map is like one natural observer field bounded by a cosmological horizon. If our hill dweller had no knowledge of the earth outside what he can see, he might decide that the earth originates at his horizon and this belief would be corroborated by the strange curvature effects that he observes in objects coming over his horizon. This belief is analogous to the belief in a big bang at the limit of our visible universe.

This analogy makes it clear that our model is very much bigger (and longer lived) than the conventional model. Indeed it could be indefinitely long-lived and of infinite size. However, as we shall see, there is evidence that the universe is bounded, at least as far as boundedness makes sense within a space-time without universal space slices or coherent time.

This paper is organised as follows. Section 2 contains basic definitions and the proof of the precise interrelation between redshift and expansion in a natural observer field. In Section 3 we cover the basic properties of de Sitter space on which our model is based and in Section 4 we prove that the time-like unit tangent flow on de Sitter space is Anosov. This use of de Sitter space is for convenience of description and is probably not essential. In Section 5 we cover rigorously the case of introducing one heavy body into de Sitter space and in Section 6 we discuss the general case. Here we cannot give a rigorous proof that a suitable metric exists, but we give instead two plausibility arguments that it does. Finally in Section 7 we make various remarks.

For more detail on the full program see the second author's web site at:
`http://msp.warwick.ac.uk/~cpr/paradigm/`
but please bear in mind that many parts of the program are still in very preliminary form.

We shall use the overall idea of this program namely that galaxies have supermassive centres which control the dynamic, and that stars in the spiral arms are moving outwards along the arms at near escape velocity. (It is this movement that maintains the shape of the arms and the long-term appearance of a galaxy.) However this paper is primarily intended to illustrate the possibility of a universe satisfying Hubble's law without overall expansion and not to describe our universe in detail.

2 Observer fields

A pseudo-Riemannian manifold L is a manifold with a non-degenerate quadratic form g on its tangent bundle called the *metric*. A *space-time* is a pseudo-Riemannian 4–manifold equipped with a metric of signature $(-,+,+,+)$. The metric is often written as ds^2, a symmetric quadratic expression in differential 1–forms. A tangent vector v is *time-like* if $g(v) < 0$, *space-like* if $g(v) > 0$ and *null* if $g(v) = 0$. The set of null vectors at a point form the *light-cone* at that point and this is a cone on two copies of S^2. A choice of one of these determines the *future* at that point and we assume *time orientability*, ie a global choice of future pointing light cones. An *observer field* on a region U in a space-time L is a smooth future-oriented time-like unit vector field on U. It is *natural* if the integral curves (field lines) are geodesic and the perpendicular 3–plane field is integrable (giving normal space slices). It then follows that the field determines a coherent notion of time: a time coordinate that is constant on the perpendicular space slices and whose difference between two space slices is the proper time along any field line; this is proved in [6, Section 5].

A natural observer field is *flat* if the normal space slices are metrically flat. In [6] we found a dual pair of spherically-symmetric natural flat observer fields for a large family of spherically-symmetric space-times including Schwarzschild and Schwarzschild-de-Sitter space-time, namely space-times which admit metrics of the form:

$$ds^2 = -Q\,dt^2 + \frac{1}{Q}\,dr^2 + r^2\,d\Omega^2 \tag{1}$$

where Q is a positive function of r. Here t is thought of as time, r as radius and $d\Omega^2$, the standard metric on the 2–sphere, is an abbreviation for $d\theta^2 + \sin^2\theta\,d\phi^2$ (or more symmetrically for $\sum_{j=1}^{3} dz_j^2$ restricted to $\sum_{j=1}^{3} z_j^2 = 1$). The Schwarzschild metric is defined by $Q = 1 - 2M/r$, the de Sitter metric by $Q = 1 - (r/a)^2$ and the combined Schwarzschild de Sitter metric by $Q = 1 - 2M/r - (r/a)^2$ (for $M/a < 1/\sqrt{27}$). Here M is mass (half the Schwarzschild radius) and a is the cosmological radius of curvature of space-time. In these cases one of these observer fields is expanding and the other contracting and it is natural to describe the expanding field as the "escape" field and the dual contracting field as the "capture" field. The expansive field for Schwarzschild-de-Sitter space-time is the main ingredient in our redshifted observer field.

Redshift in a natural observer field

The redshift z of an emitter trajectory of an observer field as seen by a receiver trajectory is given by

$$1 + z = \frac{dt_r}{dt_e} \tag{2}$$

for the 1-parameter family of null geodesics connecting the emitter to the receiver in the forward direction, where t_e and t_r are proper time along the emitter and receiver trajectories respectively.

Natural observer fields are ideally suited for a study of redshift. It is easy to define a local coefficient of expansion or contraction of space. Choose a direction in a space slice and a small interval in that direction. Use the observer field to carry this to a nearby space slice. The interval now has a possibly different length and comparing the two we read a coefficient of expansion, the relative change of length divided by the elapsed time. Intuitively we expect this to coincide with the instantaneous red or blue shift along a null geodesic in the same direction and for total red/blue shift (meaning $\log(1+z)$) to coincide with expansion/contraction integrated along the null geodesic. As this is a key point for the paper we give a formal proof of this fact.

The metric has the form

$$ds^2 = -dt^2 + g_{ij}(x,t)dx^i dx^j \tag{3}$$

where g is positive definite. The observer field is given by $\dot{x} = 0, \dot{t} = 1$. It has trajectories $x = \text{const}$ with proper time t along them.

The rate ρ of expansion of space in spatial direction ξ, along the observer field, is given by

$$\rho(\xi) = \frac{\partial}{\partial t} \log \sqrt{g_{ij}\xi^i\xi^j} = \frac{1}{2}\frac{\partial g_{ij}}{\partial t}\xi^i\xi^j/(g_{ij}\xi^i\xi^j).$$

We claim that $\log(1+z)$ can be written as $\int \rho(v)\, dt$ where $v(t)$ is the spatial direction of the velocity of the null geodesic at time t.

We prove this first for one spatial dimension. Then the null geodesics are specified by $g(\frac{dx}{dt})^2 = 1$. Without loss of generality, take $x_e < x_r$, thus $\frac{dt}{dx} = \sqrt{g}$. Differentiating with respect to initial time t_e, we obtain

$$\frac{d}{dx}\frac{dt}{dt_e} = \frac{1}{2}\frac{\partial g}{\partial t}g^{-1/2}\frac{dt}{dt_e}$$

Thus

$$\log(1+z) = \int_{x_e}^{x_r} \frac{1}{2}\frac{\partial g}{\partial t}g^{-1/2}\,dx.$$

Use $\frac{dt}{dx} = \sqrt{g}$ to change variable of integration to t:

$$\log(1+z) = \int_{t_e}^{t_r} \frac{1}{2}\frac{\partial g}{\partial t}g^{-1}\,dt$$

But $\rho(v) = \frac{1}{2}\frac{\partial g}{\partial t}g^{-1}$. So

$$\log(1+z) = \int_{t_e}^{t_r} \rho(v)\,dt.$$

To tackle the case of n spatial dimensions, geodesics of a metric G on space-time are determined by stationarity of

$$E = \int \frac{1}{2}G_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}d\lambda \tag{4}$$

over paths connecting initial to final position in space-time. This includes determination of an affine parametrisation λ. The null geodesics are those for which $E = 0$. Consider a null geodesic connecting the trajectory x_e to the trajectory x_r. Generically, it lies in a smooth 2-parameter family of geodesics connecting x_e to x_r, parametrised by the initial and final times t_e, t_r and without loss of generality with all having the same interval of affine parameter. A 1-parameter subfamily of these are null geodesics, namely those for which $E = 0$. Differentiating (4) with respect to a change in (t_e, t_r), and using stationarity of E with respect to fixed endpoints, we obtain

$$\delta E = p_0(t_r)\delta t_r - p_0(t_e)\delta t_e,$$

where

$$p_0 = G_{0\beta}\frac{dx^\beta}{d\lambda} \tag{5}$$

(0 denoting the t-component). So the subfamily for which $E = 0$ satisfies

$$\frac{dt_r}{dt_e} = \frac{p_0(t_e)}{p_0(t_r)}.$$

Thus

$$\log(1+z) = \log|p_0(t_e)| - \log|p_0(t_r)|,$$

which is minus the change in $\log|p_0|$ from emitter to receiver (note that $p_0 < 0$ for metric (3)).

Now

$$\frac{d}{dt}\log|p_0| = \frac{dp_0}{dt}p_0^{-1}.$$

Hamilton's equations for geodesics with Hamiltonian $\frac{1}{2}G^{\alpha\beta}p_\alpha p_\beta$ give

$$\frac{dp_0}{d\lambda} = -\frac{1}{2}\frac{\partial G^{\alpha\beta}}{\partial t}p_\alpha p_\beta, \ \frac{dt}{d\lambda} = G^{0\beta}p_\beta. \tag{6}$$

So

$$\frac{d}{dt}\log|p_0| = \frac{-\frac{1}{2}\frac{\partial G^{\alpha\beta}}{\partial t}p_\alpha p_\beta}{p_0 G^{0\beta}p_\beta}.$$

Now $G^{\alpha\beta}G_{\beta\gamma} = \delta^\alpha_\gamma$ implies that

$$\frac{\partial G^{\alpha\beta}}{\partial t} = -G^{\alpha\gamma}\frac{\partial G_{\gamma\delta}}{\partial t}G^{\delta\beta}$$

and $G^{\delta\beta}p_\beta = \frac{dx^\delta}{d\lambda}$ which we denote by v^δ, so

$$\frac{d}{dt}\log|p_0| = \frac{\frac{1}{2}\frac{\partial G_{\alpha\beta}}{\partial t}v^\alpha v^\beta}{p_0 v^0}.$$

For a null geodesic, $p_0 v^0 = -p_i v^i$ (where the implied sum is over only spatial components). For our form of metric (3), $-p_i v^i = -g_{ij}v^i v^j$ and the numerator also simplifies to only spatial components, so we obtain

$$\frac{d}{dt}\log|p_0| = -\frac{\frac{1}{2}\frac{\partial g_{ij}}{\partial t}v^i v^j}{g_{ij}v^i v^j} = -\rho(v). \tag{7}$$

So

$$\log(1+z) = \int_{t_e}^{t_r}\rho(v)\,dt \tag{8}$$

as desired.

Luminosity

Finally, to derive a Hubble law, we must compute the luminosity distance, ie that length d_L such that the received power per unit perpendicular area (in receiver frame) is the emitted power per unit solid angle (in emitter frame) divided by the square of $(1+z)d_L$. Some authors leave out the factor $(1+z)$, but it is natural to include it in the definition to take into account the trivial effects of redshift on received power.

For general metrics, d_L reflects focussing effects, but if one specialises to metrics satisfying Einstein's equations in vacuum (but allowing cosmological constant) then the focussing equation, [7, page 582], applied to null geodesics implies that d_L is precisely the change in affine parameter λ along the null geodesic from emitter to receiver, scaled so that $\frac{d\lambda}{dt} = 1$ in the emitter frame at the emitter.

For metrics of the form (3), we have from the inverse relation to (5), or the second of Hamilton's equations (6)

$$\frac{dt}{d\lambda} = G^{0\beta} p_\beta = |p_0|,$$

so the change in λ is

$$\int_{t_e}^{t_r} d\lambda = \int_{t_e}^{t_r} \frac{dt}{|p_0|}.$$

Hence on the vacuum assumption

$$d_L = |p_0|(t_e) \int_{t_e}^{t_r} \frac{dt}{|p_0|}.$$

Using (7) we can write

$$|p_0|(t) = |p_0|(t_e) e^{-\int_{t_e}^{t} \rho(v)(t')dt'},$$

so

$$d_L = \int_{t_e}^{t_r} e^{\int_{t_e}^{t} \rho(v)(t')dt'} dt. \tag{9}$$

Although this appears to disagree with the standard result $d_L = S(t_r) \int_{t_e}^{t_r} \frac{dt}{S(t)}$ for the special case of isotropic homogeneous space with scale factor $S(t)$ (for example [7, equation (29.7)]), our formula is derived assuming a vacuum, which restricts the overlap of the two results to the case of exponential $S(t)$, for which they give the same answer.

Compare (9) with (8), written in the form

$$z = e^{\int_{t_e}^{t_r} \rho(v)(t')dt'} - 1.$$

If $\rho(v) = \rho_o$ constant along the null geodesic then this gives

$$z = e^{\rho_o \Delta t} - 1,$$

where $\Delta t = t_r - t_e$, and (9) gives

$$d_L = \frac{1}{\rho_0}(e^{\rho_0 \Delta t} - 1).$$

Thus

$$z = \rho_0 d_L,$$

which is an exact Hubble law.

If $\rho(v)$ is not constant along the null geodesic then the relation between z and d_L is not so simple, but if $\rho(v)$ averages to a value ρ_0 along null geodesics then an approximate relation of the form $z \approx \rho_0 d_L$ is obtained.

Uniform expansion and the Schwarzschild-de-Sitter case

An important special case of the metric is when g is of the form $\lambda(t)h$ where λ is a positive function and h is independent of t. Locally this defines the warped product of a 3–manifold with time. This is the class of metrics used in conventional cosmology (the Friedman-Lemaître-Robertson-Walker or FLRW–metrics). The further special case where h is the standard quadratic form for Euclidean space $\mathbb{R}^3$ and $\lambda(t) = \exp(2t/a)$ is the unique fully homogeneous FLRW–metric. This metric is uniformly expanding, has non-zero cosmological constant (CC) namely $3/a^2$ and is the most natural choice of metric for an expanding universe.

As remarked above, the Schwarzschild-de-Sitter and Schwarzschild metrics admit flat natural observer fields. In [6, Section 6] we calculated the space expansion/contraction in the three principal directions. The average is always expansive and, as we shall see in final remark 7.4, of a size appropriate for Hubble's law.

3 de Sitter space

We give here a summary of the properties of de Sitter space that we shall need. Full proofs can be found in [3].

Definitions

Minkowski n–space $\mathbb{M}^n$ is $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ (time cross space) equipped with the standard $(-, +, \ldots, +)$ metric. The time coordinate is x_0 and the space coordinates are $x_1, \ldots, x_{n-1}$. The *Lorentz n–group* is the group of isometries of Minkowski space fixing $\mathbf{0}$ and preserving the time direction. This implies that a Lorentz transformation is a linear isomorphism of $\mathbb{R}^n$. If, in addition to preserving the time direction, we also preserve space orientation then the group can be denoted $\mathrm{SO}(1, n-1)$. Notice that a Lorentz transformation which preserves the x_0–axis is an orthogonal transformation of the perpendicular $(n-1)$–space, thus $\mathrm{SO}(n-1)$ is a subgroup of $\mathrm{SO}(1, n-1)$ and we refer to elements of this subgroup as (Euclidean) rotations about the x_0–axis.

Minkowski 4–space $\mathbb{M}$ is just called *Minkowski space* and the Lorentz 4–group is called the *Lorentz group*.

Now go up one dimension. *Hyperbolic* 4*–space* is the subset of $\mathbb{M}^5$

$$\mathbb{H}^4 = \{\|\mathbf{x}\|^2 = -a^2,\ x_0 > 0 \mid \mathbf{x} \in \mathbb{M}^5\}$$

and *de Sitter space* is the subset of $\mathbb{M}^5$

$$\mathrm{deS} = \{\|\mathbf{x}\|^2 = a^2 \mid \mathbf{x} \in \mathbb{M}^5\}.$$

There is an isometric copy $\mathbb{H}^4_-$ of hyperbolic space with $x_0 < 0$. The induced metric on hyperbolic space is Riemannian and on de Sitter space is Lorentzian. Thus de Sitter space is a space-time. The *light cone* is the subset

$$L = \{\|\mathbf{x}\| = 0 \mid \mathbf{x} \in \mathbb{M}^5\}$$

and is the cone on two 3–spheres with natural conformal structures (see below). These are S^3 and S^3_- where S^3 is in the positive time direction and S^3_- negative.

The constant a plays the role of (hyperbolic) radius and we think of it as the cosmological radius of curvature of space-time.

Points of

$$S^3 \cup S^3_- \cup \mathrm{deS} \cup \mathbb{H}^4 \cup \mathbb{H}^4_-$$

are in natural bijection with the set of half-rays from the origin and we call this *half-ray space*. $\mathrm{SO}(1, 4)$ acts on half-ray space preserving this decomposition and is easily seen to act transitively on each piece. Planes through the origin meet half-ray space in *lines* which come in three types: *time-like* (meeting $\mathbb{H}^4$), *light-like* (tangent to light-cone) and *space-like* (disjoint from the light cone). Symmetry considerations show that lines meet $\mathbb{H}^4$ and deS in geodesics (and all geodesics are of this form). Figure 1 is a projective picture illustrating these types.

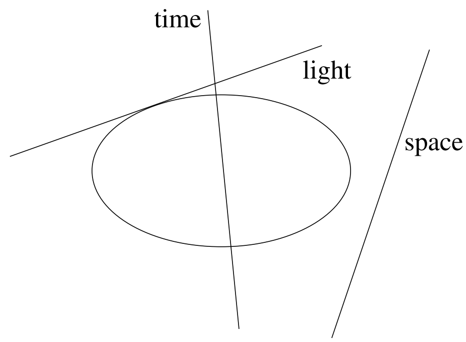

Figure 1: Types of lines (geodesics)

$\mathbb{H}^4$ with the action of $\mathrm{SO}(1,4)$ is the *Klein model* of hyperbolic 4–space. S^3 is then the sphere at infinity and $\mathrm{SO}(1,4)$ acts by conformal transformations of S^3 and indeed is isomorphic to the group of such transformations.

$\mathrm{SO}(1,4)$ also acts as the group of time and space orientation preserving isometries of deS and can be called the *de Sitter group* as a result.

A simple combination of elementary motions of deS proves that $\mathrm{SO}(1,4)$ acts transitively on lines/geodesics of the same type in half-ray space and indeed acts transitively on *pointed lines*. In other words:

Proposition 1 *Given geodesics l, m of the same type and points $P \in l$, $Q \in m$, there is an isometry carrying l to m and P to Q.*

It is worth remarking that topologically deS is $\mathbb{R} \times S^3$. Geometrically it is a hyperboloid of one sheet ruled by lines and each tangent plane to the light cone meets deS in two ruling lines. These lines are light-lines in $\mathbb{M}^5$ and hence in deS. (See Figure 2, which is taken from Moschella [8].)

The expansive metric

Let Π be the 4–dimensional hyperplane $x_0 + x_4 = 0$. This cuts deS into two identical regions. Concentrate on the upper region Exp defined by $x_0 + x_4 > 0$. Π is tangent to both spheres at infinity S^3 and S^3_-. Name the points of tangency as P on S^3 and P_- on S^3_-. The hyperplanes parallel to Π, given by $x_0 + x_4 = k$ for $k > 0$, are also all tangent to S^3 and S^3_- at P, P_- and foliate Exp by paraboloids. Denote this foliation by $\mathcal{F}$. We

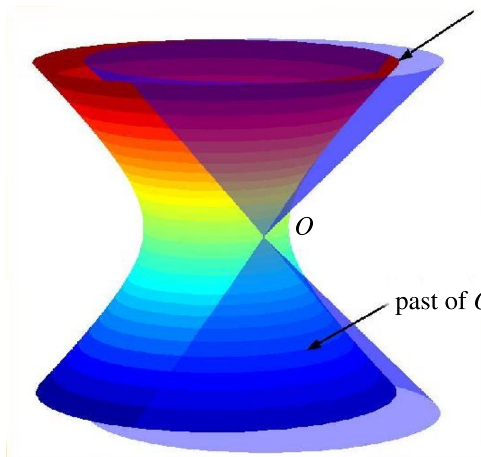

Figure 2: Light-cones: the light-cone in deS is the cone on a 0–sphere (two points) in the dimension illustrated, in fact it is the cone on a 2–sphere. The figure is reproduced with permission from [8].

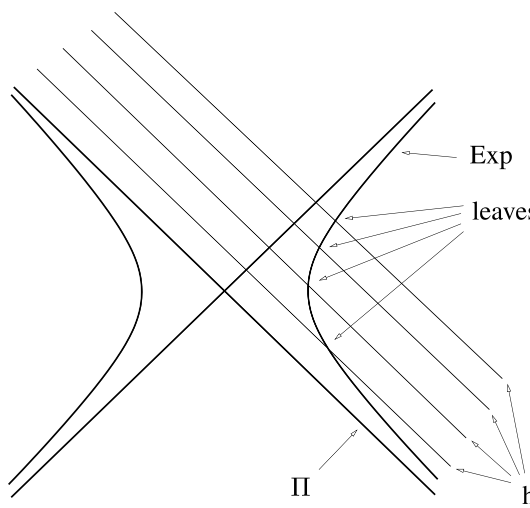

Figure 3: The foliation $\mathcal{F}$ in the (x_0, x_4)–plane

shall see that each leaf of $\mathcal{F}$ is in fact isometric to $\mathbb{R}^3$. There is a transverse foliation $\mathcal{T}$ by the time-like geodesics passing through P and P_-.

These foliations are illustrated in Figures 3 and 4. Figure 3 is the slice by the (x_0, x_4)–coordinate plane and Figure 4 (the left-hand figure) shows the view from the x_4–axis in 3–dimensional Minkowski space (2–dimensional de Sitter space). This figure and its companion are again taken from Moschella [8].

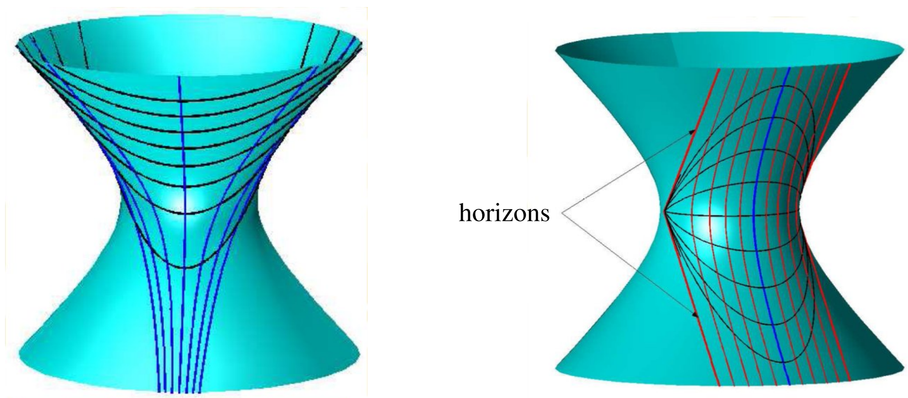

Figure 4: Two figures reproduced with permission from [8]. The left-hand figure shows the foliation $\mathcal{F}$ (black lines) and the transverse foliation $\mathcal{T}$ by geodesics (blue lines). The righthand figure shows the de Sitter metric as a subset of deS.

Let G be the subgroup of the Lorentz group which fixes P (and hence P_- and Π). G acts on Exp. It preserves both foliations: for the second foliation this is obvious, but all Lorentz transformations are affine and hence carry parallel hyperplanes to parallel hyperplanes; this proves that it preserves the first foliation. Furthermore affine considerations also imply that it acts on the set of leaves of $\mathcal{F}$ by scaling from the origin. Compare this action with the conformal action of G on S^3 (the light sphere at infinity). Here G acts by conformal isomorphisms fixing P which are similarity transformations of $S^3 - P \cong \mathbb{R}^3$. The action on the set of leaves of $\mathcal{F}$ corresponds to dilations of $\mathbb{R}^3$ and the action on a particular leaf corresponds to isometries of $\mathbb{R}^3$. By dimension considerations this gives the full group of isometries of each leaf. It follows that each leaf has a flat Euclidean metric.

We can now see that the metric on Exp is the same as the FLRW metric for a uniformly expanding infinite universe described in Section 2 above. The transverse foliation by time-like geodesics determines the standard observer field and the distance between hyperplanes defining $\mathcal{F}$ gives a logarithmic measure of time. Explicit coordinates are given in [3]. Notice that we have proved that every isometry of Exp is induced by an isometry of deS.

It is worth remarking that exactly the same analysis can be carried out for $\mathbb{H}^4$ where the leaves of the foliation given by the same set of hyperplanes are again Euclidean. This gives the usual "half-space" model for hyperbolic geometry with Euclidean horizontal sections and vertical dilation.

Time-like geodesics in Exp

We have a family of time-like geodesics built in to Exp namely the observer field mentioned above. These geodesics are all *stationary*, in the sense that they are at rest with respect to the observer field. They are all equivalent by a symmetry of Exp because we can use a Euclidean motion to move any point of one leaf into any other point. Other time-like geodesics are *non-stationary*. Here is a perhaps surprising fact:

Proposition 2 *Let l, m be any two non-stationary geodesics in* Exp*. Then there is an element of $G = \mathrm{Isom}(\mathrm{Exp})$ carrying l to m.*

Thus there in no concept of conserved velocity of a geodesic with respect to the standard observer field in the expansive metric. This fact is important for the analysis of black holes in de Sitter space, see below.

The proof is easy if one thinks in terms of hyperbolic geometry. Time-like geodesics in Exp are in bijection with geodesics in $\mathbb{H}^4$ since both correspond to 2–planes through the origin which meet $\mathbb{H}^4$. But if we use the upper half-space picture for $\mathbb{H}^4$, stationary means vertical and two non-stationary geodesics are represented by semi-circles perpendicular to the boundary. Then there is a conformal map of this boundary (ie a similarity transformation) carrying any two points to any two others: translate to make one point coincide and then dilate and rotate to get the other ones to coincide.

The de Sitter metric

There is another standard metric inside de Sitter space, namely that of form (1) with $Q = 1 - (r/a)^2$. It is essentially the metric which de Sitter himself used (change variable from de Sitter's r in (4B) of [11] to $r' = R\sin(r/R)$ and reverse the sign). The metric is illustrated in Figure 4 on the right. This metric is static, in other words there is a time-like Killing vector field (one whose associated flow is an isometry). The region where it is defined is the intersection of $x_0 + x_4 > 0$ defining Exp with $x_0 - x_4 < 0$ (defining the reflection of Exp in Ξ, the (x_1, x_2, x_3, x_4)–coordinate hyperplane). The observer field, given by the Killing vector field, has exactly one geodesic leaf, namely the central (blue) geodesic. The other leaves (red) are intersections with parallel planes not passing through the origin. There are two families of symmetries of this subset: an SO(3)–family of rotations about the central geodesic and shear along this geodesic (in the (x_0, x_1)–plane). Both are induced by isometries of deS.

This metric accurately describes the middle distance neighbourhood of a black hole in empty space with a non-zero CC. The embedding in deS is determined by the choice of central time-like geodesic. Proposition 2 then implies that there are precisely two types of black hole in a standard uniformly expanding universe. There are stationary black holes which, looking backwards in time, all originate from the same point. Since nothing real is ever completely at rest, this type of black hole is not physically meaningful. The second class are black holes with non-zero velocity. Looking backwards in time these all come from "outside the universe" with infinite velocity (infinite blueshift) and gradually slow down to asymptotically zero velocity. If this description has any relation to the real universe then this phenomenon might give an explanation for observed gamma ray bursts. In any case it underlines clearly the unreality of assuming the existence a standard uniformly expanding universe containing black holes.

In the next section we consider a metric which accurately describes the immediate neighbourhood of the black hole as well as its surroundings.

The contractive metric

Reflecting in Ξ (the (x_1, x_2, x_3, x_4)–coordinate hyperplane) carries Exp to the subset Cont defined by $x_0 - x_4 < 0$. Exactly the same analysis shows that Cont has an FLRW-metric with constant warping function $\exp(-2t/a)$, which corresponds to a uniformly *contracting* universe with *blueshift* growing linearly with distance. Since the subsets Exp and Cont overlap (in the region where the de Sitter metric is defined), by homogeneity, any small open set in deS can be given two coordinate systems, one of which corresponds to the uniformly expanding FLRW-metric and the other to the uniformly contracting FLRW-metric. These overlapping coordinate systems can be used to prove that the time-like geodesic flow is Anosov.

4 The Anosov property

A C^1 flow ϕ on a manifold M is equivalent to a vector field v by $d\phi_\tau(x) = v(\phi_\tau(x))$ where ϕ_τ is the diffeomorphism given by flowing for time τ. The flow is *Anosov* if there is a splitting of the tangent bundle TM as a direct sum of invariant subbundles $E^- \oplus E^+ \oplus \mathbb{R}v$ such that, with respect to a norm on tangent vectors, there are real numbers $C, \lambda > 0$ such that $u \in E^-$ respectively E^+ implies $|u_t| \leq C\exp(-\lambda|t|)\,|u|$ for all $t > 0$ respectively $t < 0$ where u_t is $d\phi_t(u)$. If M is compact then different norms do not affect the Anosov property, only the value of C, but if M is non-compact one must specify a norm.

The *time-like geodesic flow* on a space-time is the flow on the negative unit tangent bundle $T_{-1}(M) = \{v \in TM \mid g(v) = -1\}$ induced by flowing along geodesics. More geometrically we can think of $W = T_{-1}(M)$ as the space of germs of time-like geodesics. Thus a point in W is a pair (X, x) where $X \in M$ and x is an equivalence class of oriented time-like geodesics through X, where two are equivalent if they agree near X; this is obviously the same as specifiying a time-like tangent vector at X up to a positive scale factor. The geodesic flow ψ is defined by $\psi_\tau(X, x) = x(\tau)$ where $x()$ is the geodesic determined by x parametrised by distance from X.

Proposition 3 $T_{-1}(\mathrm{deS})$ *is a Riemannian manifold and* ψ *is Anosov.*

To prove the proposition we need to specify the norm on the tangent bundle of $W = T_{-1}(\mathrm{deS})$ and prove the Anosov property. Using Proposition 1 we need only do this at one particular point (X, x) in W and then carry the norm (and the Anosov parameters) around W using isometries of deS. We choose to do this at (Q, g) where $Q = (0, 0, 0, a)$ and g is determined by the (x_0, x_4)–plane. These are the central point and vertical geodesic in Figure 4. Recall that Ξ is the hyperplane orthogonal to g at Q (orthogonal in either Minkowski or Euclidean metric is the same here!) and let Ξ' be a nearby parallel hyperplane. A geodesic near to g at Q can be specified by choosing points T, T' in Ξ, Ξ' near to Q and to specify a point of W near to (Q, g) we also need a point on one of these geodesics and we can parametrize such points by hyperplanes parallel and close to Ξ. We can identify the tangent space to W at (Q, g) with these nearby points of W in the usual way and for coordinates we have Euclidean coordinates in Ξ and Ξ' and the distance between hyperplanes. This gives us a positive definite norm on this tangent space.

Now recall that we have a foliation $\mathcal{T}$ of deS near Q by time-like geodesics passing through P and P_- (the time curves in the expanding metric) and dually (reflecting in Ξ) another foliation $\mathcal{T}'$ which are the time curves in the contracting metric ie geodesics passing through P' and P'_- the reflected points. Let E^+ be the subspace determined by $\mathcal{T}$ at points of Ξ near Q and let E^- be determined similarly by $\mathcal{T}'$. These meet at (Q, g) and span a subspace of codimension 1. The remaining 1–dimensional space is defined by (Q_ϵ, g) where Q_ϵ varies through points of g near Q. The Anosov property holds with $\lambda = a$ and $C = 1$ by the expanding and contracting properties of the two metrics which we saw above.

Alternatively, one can study the Jacobi equation $v'' = Mv = -R(u, v, u)$ for linearised perpendicular displacement v to a time-like geodesic with tangent u (without loss of generality, unit length). Now $\mathrm{Tr}M = -\mathrm{Ric}(u,u) = -\Lambda g(u,u) = \Lambda$. But de Sitter space has rotational symmetry about any time-like vector, in particular u, so M is a multiple of the identity, hence $\frac{\Lambda}{3}$. Thus $v'' = \frac{\Lambda}{3}v$ and $v(t) = v^+e^{-t/a} + v^-e^{t/a}$, demonstrating the splitting into vectors which contract exponentially in forwards and backwards time respectively.

5 The Schwarzschild de Sitter metric

In this section we look at the effect of introducing a black hole into de Sitter space. There is an explicit metric which modifies the standard Schwarzschild metric to be valid in space-time with a CC (see eg Giblin–Marolf–Garvey [4, Equation 3.2]) given by:

$$ds^2 = -Q(r)\,dt^2 + Q(r)^{-1}\,dr^2 + r^2\,d\Omega^2 \tag{10}$$

Here $Q(r) = 1 - 2M/r - (r/a)^2$ where M is mass (half the Schwarzschild radius) and the CC is $3/a^2$ as usual.

There are several comments that need to be made about this metric.

(A) It is singular where $Q = 0$ ie when $r = 2M$ approx and when $r = a$ approx (we are assuming that a is large). The singularity at $r = 2M$ is at (roughly) the usual Schwarzschild radius. It is well known to be removable eg by using Eddington–Finkelstein coordinates. The singularity at $r = a$ is a "cosmological horizon" in these coordinates and is again removable. This is proved in [4] and more detail is given below.

(B) The metric is static in the sense that the time coordinate corresponds to a time-like Killing vector field. But like the de Sitter metric described above, the time field is completely unnatural so this stasis has no physical meaning. Indeed for these coordinates *no* time curve is a geodesic.

(C) Comparing equations (1) in the de Sitter case ($Q = 1 - (r/a)^2$) and (10) we see that the former is precisely the same as the latter if the Schwarzschild term $-2M/r$ is removed. In other words if the central mass (the black hole) is removed the Schwarzschild metric (modified for CC) becomes the de Sitter metric. Physically what this means is that as r increases (and the Schwarzschild term tends to zero) the metric approximates the de Sitter metric closely. Thus the metric embeds the black hole metric with CC into de Sitter space and in particular we can extend the metric past the cosmological horizon at $r = a$ which can now be seen as a removable singularity. This will be proved rigorously below.

(D) We can picture the metric inside deS by using Figure 4 (right). Imagine that the central geodesic (blue) is a thick black line. This is the black hole. The other vertical (red) lines are the unnatural observer field given by the time lines. The horizontal black curved lines are the unnatural space slices corresponding to the unnatural observer field. The thick red lines marked "horizon" are the "cosmological horizon" which is removable and we now establish this fact rigorously.

Extending the expansive field to infinity

We shall use the ideas of Giblin et al [4] to extend the metric (10) beyond the horizon at $r = a$ to the whole of Exp. In [6] we proved that there is a unique pair of flat space slices for a spherically-symmetric space-time which admits such slices. Since this is true for both the Schwarzschild-de-Sitter metric and Exp it follows that the expansive observer field (the escape field) found in [6] for the Schwarzschild-de-Sitter metric merges into the standard expansive field on Exp. We shall need to make a small change to de Sitter space to achieve the extension but we shall not disturb the fact that the space satisfies Einstein's equation Ric $= \Lambda g$ with CC $\Lambda = 3/a^2$. Recall that $Q(r) = 1 - 2M/r - (r/a)^2$.

A *horizon* is an r_0 for which $Q(r_0) = 0$. Q has an inner horizon r_b and an outer horizon r_c. It is convenient to treat r_b, r_c as the parameters instead of a, M. Thus

$$Q(r) = -\frac{1}{a^2 r}(r - r_b)(r - r_c)(r + r_b + r_c),$$

and then a, M become functions of r_b, r_c, as

$$a^2 = r_b^2 + r_c^2 + r_b r_c, M = r_b r_c (r_b + r_c)/2a^2.$$

Suppose $\kappa := \frac{1}{2}Q'(r_0) \neq 0$ and $Q''(r_0) < 0$. This is true at both horizons, but we shall just extend beyond r_c.

$$Q'(r) = \frac{2}{a^2 r^2}(a^2 M - r^3)$$

so $\kappa = (a^2 M - r_c^3)/a^2 r_c^2$.

We will achieve the goal as a submanifold of $\mathbb{R}^5$ with coordinates (X, T, y_1, y_2, y_3)

$$X^2 - T^2 = Q(r)/\kappa^2.$$

The submanifold is given the metric

$$ds^2 = -dT^2 + dX^2 + \beta(r)dr^2 + r^2 d\Omega^2$$

with

$$\beta(r) = \frac{4\kappa^2 - Q'(r)^2}{4\kappa^2 Q(r)}.$$

The numerator of β contains a factor $(r - r_c)$ which cancels the factor $(r - r_c)$ of Q, and $\beta(r) \in (0, \infty)$ for all $r \in (r_b, \infty)$.

The above submanifold and metric are obtained using the coordinate change

$$X(r,t) = \frac{1}{\kappa}\sqrt{Q(r)}\cosh(\kappa t),$$

$$T(r,t) = \frac{1}{\kappa}\sqrt{Q(r)}\sinh(\kappa t).$$

The metric satisfies $\mathrm{Ric} = \Lambda g$ because the original metric does in the region $Q(r) > 0$, so the new one does for $r \in (r_b, r_c)$, but the submanifold and the coefficients of the metric are given by rational expressions in r, so $\mathrm{Ric} = \Lambda g$ wherever g is regular.

Asymptotic form for large r

$$\beta(r) = \frac{a^2((r^3 - a^2M)^2 r_c^4 - (r_c^3 - a^2M)^2 r^4)}{(r_c^3 - a^2M)^2 r^3 (r^3 - a^2 r + 2Ma^2)}$$

It is convenient to use $r_c^3 - a^2M = \frac{1}{2} r_c (3r_c^2 - a^2)$. Then:

$$\beta(r) = \frac{4a^2 r_c^2}{(3r_c^2 - a^2)^2} \frac{1 - \frac{(3r_c^2 - a^2)^2}{4r_c^2} r^{-2} - 2a^2 M r^{-3} + a^4 M^2 r^{-6}}{1 - a^2 r^{-2} + 2Ma^2 r^{-3}}.$$

The factor at the front can be written $(\kappa a)^{-2}$, and is close to 1 for M small, so we write $\beta(r) = (\kappa a)^{-2}\tilde{\beta}(r)$. Now $(3r_c^2 - a^2)/2r_c$ is close to a for small M, so it is sensible to extract the leading correction $-a^2 r^{-2}$ in the numerator, writing

$$4a^2 r_c^2 - (3r_c^2 - a^2)^2 = r_b(8r_c^3 + 7r_b r_c^2 - 2r_b^2 r_c - r_b^3).$$

Thus:

$$\tilde{\beta}(r) = \frac{1 - a^2 r^{-2} + \frac{r_b}{4r_c^2}(8r_c^3 + 7r_b r_c^2 - 2r_b^2 r_c - r_b^3) r^{-2} - 2a^2 M r^{-3} + a^4 M^2 r^{-6}}{1 - a^2 r^{-2} + 2Ma^2 r^{-3}}$$

$$= 1 + \frac{r_b(8r_c^3 + 7r_b r_c^2 - 2r_b^2 r_c - r_b^3)}{4r_c^2(r^2 - a^2)} - \frac{4Ma^2}{r(r^2 - a^2)} + O(M^2 a^3 r^{-5}).$$

An an alternative we can scale X, T to $X' = \kappa a X, T' = \kappa a T$ so that the submanifold is

$$X'^2 - T'^2 = a^2 Q(r) = a^2 - r^2 - 2Ma^2/r$$

which is an $O(Ma^2/r^3)$ perturbation of the standard hyperboloid for large r. The metric then is

$$ds^2 = (\kappa a)^{-2}(-dT'^2 + dX'^2 + \tilde{\beta}(r)dr^2) + r^2 d\Omega^2.$$

Interestingly, this is not asymptotic to the de Sitter metric as $r \to \infty$ because of the factor $(\kappa a)^{-2}$, although it does tend to 1 as $M \to 0$. Scaling the metric by this factor does not help because then the last term becomes $(\kappa a)^2 r^2 d\Omega^2$.

6 The model

Recall that we are assuming that most galaxies have supermassive centres and that most stars in those galaxies are travelling on near escape orbits. This implies that most of the sources of light from distant galaxies is from stars fitting into an escape field. We also fit into the escape field from our own galaxy. In [6] we proved that these fields are expansive, see Section 2, and as we saw above, they each individually merge into the standard expansive field on de Sitter space. We shall give two plausibility arguments to show that this extension works with many black holes and therefore for many black holes in de Sitter space there is locally a (roughly uniformly) expansive field outside the black holes which we call the *dominant* observer field. Our model is de Sitter space with many black holes, one for each galaxy in the universe, and anywhere in the model there is a dominant observer field which locally contains all the information from which expansion (and incorrectly, the big bang) are deduced. Thus any observer in any of these fields observes a pattern of uniform redshift and Hubble's law.

There are dual contractive fields and the two are in balance, so that this universe "in the large" is not expanding. Indeed the expansive field only covers the visible parts of the universe. Most of the matter in the universe is contained in the supermassive black holes at the centres of galaxies and this matter is not in any sense in an expanding universe. The space slices in De Sitter space are infinite but there is no need to assume this for the real universe (merely that they are rather large) for a description along these lines to make sense; cf final remark 7.2.

The first plausibility argument: Anosov property

We saw that de Sitter space has an Anosov property. Anosov dynamical systems are structurally stable. Even more, for any set of time-like geodesics, if weak interactions are added between the world-lines there is a consistent set of deformed world-lines which remain uniformly close to the original ones. Thus if the dynamics of black holes in de Sitter space were just of the form of an interaction force between world-lines then the Anosov property would allow us to construct solutions as small perturbations of arbitrary sets of sufficiently separated time-like geodesics, namely whose pairwise closest approaches exceed $(Ma^2)^{1/3}$, where M is the reduced mass for the pair (ie $1/M = 1/M_1 + 1/M_2$). General relativity is more complicated than this, the interaction being mediated by a metric. In particular, interaction of black holes could produce gravitational waves which would not be captured by the above scenario. Nevertheless this argument strongly suggests that there are space-times with $\mathrm{Ric} = \Lambda g$ containing arbitrarily many (but well separated) interacting black holes, to which our construction of expanding observer fields and hence our conclusions about redshift would apply.

The second plausibility argument: the charged case

Kastor and Traschen [5] give an explicit solution for the metric appropriate to many charged black holes in de Sitter space. Again this proof does not apply to the real universe because the assumption made is that the charge cancels out gravitational attraction whereas well known observations (eg Zwicky [12]) clearly show that gravity is alive and well for galaxies in the real universe.

Nevertheless it does again strongly suggest that the result we want is correct.

7 Final remarks

7.1 Evidence from the Hubble ultra deep field

Note that the expansion from a heavy centre is not uniform. Near the centre the radial coordinate contracts and the tangential coordinates expand, with a nett effect that is still expansive on average. This is described in detail in [6, Section 6]. When these expansive fields are fitted together to form the global expansive field for the universe, because of the non-uniformity in each piece, the result is highly non-uniform in other words is filled with gravitational waves. There are also very probably many other sources of gravitational waves which add to the non-uniformity.

There is strong direct evidence for gravitational waves in deep observations made by the Hubble telescope, see [9]. It is also worth commenting that there may also be indirect evidence provided by gamma ray bursts which may well be caustics in the developing observer field caused by the same non-uniformity, but see also the comments about black holes entering an expanding universe made in Section 3 above.

7.2 Bounded space slices

The model we constructed above has infinite space slices for the dominant observer field, based as it is on the flat slicing of de Sitter space. There is however a nearby model where the slicing is tilted a little so that the space slices are bounded. This has the effect of causing the expansion to drop off over very long distances and we cannot resist pointing out that this is exactly what is observed [1, 2] and conventionally ascribed to dark energy.

7.3 Cosmological constant

We used de Sitter space as a substrate for our model because of the wealth of structure and results associated with this space. It has the consequence that we are assuming a non-zero CC. However it is quite plausible that one could splice together many copies of Schwarzschild space, which has zero CC, and obtain a similar model but here we have no hint at any rigorous proof.

7.4 Estimates

Our attitude throughout this paper has been the construction of a model which demonstrates that redshift and expansion do not have to be associated with a big bang and therefore we have avoided direct estimates of the size of the expansion that we have considered. But we finish with just this. The estimate of linear expansion for the natural flat observer field near a black hole of mass M is $T/2$ where $T = \sqrt{2M}r^{-3/2}$, [6, Section 5]. Further out the rate tends towards an expansion rate of $1/a$ corresponding to the CC. Galaxies are typically spaced about 10^7 light years apart. So we expect an expansion of about $\sqrt{M} \times 10^{-10.5}$ and if this is the observed Hubble expansion of $10^{-10.5}$ then we find M is of the order of one light year. In solar masses this is $10^{12.5}$ solar masses. This is within the expected range for a galactic centre though we should point out that it is 2.5 orders of magnitude smaller than the estimate of 10^{15} solar masses obtained in the main paper in this program [10]. This discrepancy is probably due to ignoring angular momentum, see the next remark.

7.5 Health warning

This paper has used a fully relativistic treatment throughout. However the main paper in the program [10] uses a combination of Newtonian dynamics and relativistic effects especially in the derivation of the rotation curve of a galaxy and the full dynamic. It needs rewriting in a fully relativistic form using the Kerr metric because non-zero angular momentum is crucial to the treatment. This probably means that the analysis in this paper will need to be changed.

References

[1] *The supernova cosmology project*, Lawrence Berkeley Laboratory, see http://panisse.lbl.gov

[2] *The high-Z supernova search*, see http://cfa-www.harvard.edu/cfa/oir/Research/supernova/

[3] *Notes on de Sitter space*, http://msp.warwick.ac.uk/~cpr/paradigm/notes.pdf

[4] **J T Giblin**, **D Marolf**, **R Garvey**, *Spacetime embedding diagrams for spherically symmetric black holes*, General Relativity and Gravitation 36 (2004) 83–99

[5] **D Kastor**, **J Traschen**, *Cosmological multi-black-hole solutions*, Phys Rev D 47 (1993) 5370–5

[6] **R S MacKay**, **C Rourke**, *Natural flat observer fields in spherically-symmetric space-times*, submitted for publication, available at http://msp.warwick.ac.uk/~cpr/paradigm/escape-submit.pdf

[7] **C W Misner**, **K S Thorne**, **J A Wheeler**, *Gravitation*, Freeman (1973)

[8] **U Moschella**, The de Sitter and anti-de Sitter sightseeing tour, Séminaire Poincaé 1 (2005) 1–12, available at http://www.bourbaphy.fr/moschella.pdf

[9] **C Rourke**, *Optical distortion in the Hubble Ultra-Deep Field*, available at http://msp.warwick.ac.uk/~cpr/paradigm/HUDF-2.pdf

[10] **C Rourke**, *A new paradigm for the structure of galaxies*, available at http://msp.warwick.ac.uk/~cpr/paradigm/struc-gal.pdf

[11] **W de Sitter**, *On the curvature of space*, Proc KNAW 20 (1918) 229–243

[12] **F Zwicky**, *Die Rotverschiebung von extragalaktischen Nebeln*, Helvetica Physica Acta 6 (1933) 110–127. See also **F Zwicky**, *On the Masses of Nebulae and of Clusters of Nebulae*, Astrophysical J 86 (1937) 217 DOI:10.1086/143864

Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

R.S.MacKay@warwick.ac.uk, cpr@msp.warwick.ac.uk

http://www.warwick.ac.uk/~mackay/, http://msp.warwick.ac.uk/~cpr