Palestine Journal of Mathematics

Vol. 5(Special Issue: 1) (2016) , 175–197

Are gamma-ray bursts optical illusions?

Robert S MacKay, Colin Rourke

Communicated by Ayman Badawi

MSC 2010 Classifications: 83F05, 83C20, 83B05, 83C75.

Keywords and phrases: gamma-ray burst, kinematic effect, space-time, geodesics, de Sitter space.

Abstract. We propose that a gamma-ray burst is a kinematic effect corresponding to our first entry into the region of space-time illuminated by a continuously emitting object. We illustrate this by an analysis of light rays between time-like geodesics in de Sitter space. In the two-parameter space of pairs of time-like geodesics modulo isometries we find regimes which give light curves similar to observations. We also make remarks about more general models.

1 Introduction

Gamma-ray bursts (GRBs) are intense flashes of electromagnetic radiation of cosmic origin lasting from ten milliseconds to several minutes, whose frequency starts in the gamma range and descends. They were first observed in 1967 by satellites designed to verify the nuclear test ban treaty and are now detected regularly, e.g. [BAT]. Most proposed explanations involve cataclysmic events (for a review, see [Mes]), but there is still no generally accepted mechanism.

We propose that a gamma-ray burst is a simple kinematic effect, namely the effect of our entry into the union of the forward light cones of a continuously emitting object (Weyl's "range of influence" [We30]); thus no cataclysm is required. We illustrate this proposal by calculation of the light received from a steadily emitting time-like geodesic in de Sitter space by a receiver time-like geodesic and find a parameter regime which gives light curves similar to observations. We do not propose that this is an accurate model of what is happening in the universe, but use de Sitter space as a relatively easy context in which to demonstrate the effect. At the end of the paper we make some suggestions how the effect transfers to more accurate models.

For a pair of time-like geodesics in de Sitter space in general position, there is a first receiver time t^* at which the emitter becomes visible. The received flux starts infinitely blue-shifted and infinitely large, with a non-integrable singularity. Specifically, the received flux at receiver time t is asymptotically proportional to $(t-t^*)^{-2}$ as t decreases to t^*, and it is blue-shifted with ratio ω_r/ω_e of received to emitted frequency asymptotically $1/(t-t^*)$. The infinity of received flux is regularised if the emitter intensity is a suitable integrable function of emitter time (precisely if $\int_{-\infty}^{u} e^{-u}P(u)\,du < \infty$ where $P(u)$ is the emitter power per unit solid angle at emitter time u), but nevertheless the event of first sight of an emitter is sudden, extraordinarily bright and highly blue-shifted: a gamma-ray burst.

At a certain time t_0 the received flux passes from being blue-shifted to redshifted. As $t \to +\infty$ the received flux goes to zero and the redshift goes to infinity exponentially. The duration $t_B = t_0 - t^*$ of blueshift depends on the relative disposition of the pair of geodesics. There is a two-parameter space of pairs modulo isometry. Although in most of the parameter space t_B is of the order of the de Sitter radius (which one should think of as about 12 Gyr, based on current estimates of the cosmological constant) there is a parameter regime in which $t_B \to 0$ and the received flux is especially strong. Furthermore, on the hypothesis of uniform distribution of emitters, the density for t_B goes to infinity like t_B^{-5} as $t_B \to 0$. These are our proposed reasons for why the observed GRBs are short.

The light curves we find for de Sitter space are simpler than those observed. Towards the end of the paper we explain how these more complicated curves may be fitted by extending to a more accurate model. There are two significant effects. Firstly the emitter is almost certainly not emitting steadily and the effect of variations is to modulate the light curve sharply over a very short period of receiver time. Secondly the light is likely to reach us by multiple paths owing to the presence of nearby masses or gravitational waves, which cause caustics in our backward

light cone and result in echo effects.

There are claims to measure redshifts for gamma ray bursts (e.g. [GP]; see also the website [Gr] which documents evidence for localisation of GRBs, including redshifts for around 10% of them) but we suspect that they are artefacts, for example absorption lines from an intermediate cloud.

2 De Sitter space

De Sitter space (denoted deS) is the Lorentzian manifold (pseudo-Riemannian of signature $(-,+++)$) given by restricting 5–dimensional Minkowski space M^5 with metric

$$ds^2 = -dx_0^2 + \sum_{i=1}^{4} dx_i^2$$

to the hyperboloid

$$-x_0^2 + \sum_{i=1}^{4} x_i^2 = a^2$$

for some $a > 0$ which we call the cosmological or de Sitter radius. De Sitter space, or more precisely a patch of it, was introduced in [deS17] (with a denoted by R, but we reserve R for the Ricci tensor) and proposed by Levi-Civita around the same time [LC]. The above formulation appears in Weyl's paper [We]. It is a vacuum solution of Einstein's equations with positive cosmological constant $\Lambda = 3/a^2$. Background results on de Sitter space can be found in [N].

In this section, the redshift and intensity of emission received from a time-like emitter geodesic by a time-like receiver geodesic are calculated. We imagine this must have been done long ago, but have not found a reference (though Weyl calculated redshift for a past asymptotic field of geodesics [We], and de Sitter space became a very popular subject of study in the 1950s and 1960s, e.g. Schrödinger's book [S]). De Sitter himself addressed this problem in Section 7 of [deS18], claiming redshift for emitter-receiver pairs from some assumed cosmic velocity field but without an explicit calculation.

For simplicity we scale so that $a = 1$. This is the same as using the de Sitter radius as unit for time.

By an isometry of de Sitter space, any time-like geodesic can be transformed to any other. So without loss of generality, the receiver geodesic r can be chosen to have

$$r_0(t) = \sinh(t), \quad r_1(t) = \cosh(t), \quad r_i(t) = 0 \ \text{ for } \ i = 2, 3, 4, \tag{2.1}$$

with t being proper time along it.

Similarly, the emitter geodesic $e = Pr$, for some orientation and future-preserving isometry P of M^5, which we represent by a 5×5 matrix (also denoted P). So

$$\begin{aligned} e_0(u) &= \alpha r_0(u) + \beta r_1(u) = \alpha \sinh(u) + \beta \cosh(u) \\ e_1(u) &= \gamma r_0(u) + \delta r_1(u) \ = \gamma \sinh(u) + \delta \cosh(u) \end{aligned}$$

parametrised by its proper time u, where $Q = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is the top left corner of P.

2.1 Null geodesics in deS

The null geodesics of de Sitter space are generators of the hyperboloid, i.e. straight lines in M^5 which lie in the hyperboloid, and are given by $x(\lambda) = b + \lambda v$ with $\langle b, b\rangle = 1$, $\langle v, v\rangle = 0$, $\langle b, v\rangle = 0$, with respect to the pseudo-inner-product on M^5 (recall that we have scaled so that the hyperboloid is given by $\langle b, b\rangle = 1$). There is a connecting null geodesic between two points x, y of de Sitter space if and only if $\langle x, y\rangle = 1$. (This follows from standard projective geometry: two points lying on a quadric are conjugate with respect to the quadric iff the whole line joining them lies in the quadric. It can also be quickly proved by writing $x = y + v$ and expanding $\langle x, x\rangle = 1$

using $\langle y, y\rangle = 1$, $\langle v, v\rangle = 0$.)

Thus there is a null geodesic connecting emitter to receiver if and only if

$$-(\alpha \sinh(u) + \beta \cosh(u)) \sinh(t) + (\gamma \sinh(u) + \delta \cosh(u)) \cosh(t) = 1. \tag{2.2}$$

Writing $T = e^t$, $U = e^u$, this can be expressed as

$$-AUT + BT/U + CU/T - D/(TU) = 2, \tag{2.3}$$

where

$$\begin{aligned} 2A &= \alpha + \beta - \gamma - \delta \\ 2B &= \alpha - \beta - \gamma + \delta \\ 2C &= \alpha + \beta + \gamma + \delta \\ 2D &= \alpha - \beta + \gamma - \delta. \end{aligned}$$

2.2 Conditions on Q

From (2.2) the condition for a connecting null geodesic depends on only the top left corner Q of P, and we shall see shortly that, up to an isometry fixing r, Q determines $e = P(r)$.

There are conditions on a (2×2)–matrix Q for it to be the top left corner of a future-preserving isometry matrix P. The first column must extend to a time-like unit vector, which happens iff

$$\alpha^2 - \gamma^2 \geq 1. \tag{2.4}$$

Future-preserving implies that $\alpha > 0$. There is a third condition: we need to be able to extend the columns of Q to orthogonal vectors v, w of unit length (v time-like and w space-like). By suitable rotations fixing r we can assume that $v_4 = v_5 = w_5 = 0$ so we have to find ξ, η, ζ so that

$$-\alpha^2 + \gamma^2 + \xi^2 = -1 \tag{2.5}$$

$$-\beta^2 + \delta^2 + \eta^2 + \zeta^2 = 1 \tag{2.6}$$

$$-\alpha\beta + \gamma\delta + \xi\eta = 0. \tag{2.7}$$

Then the usual argument for completing an orthonormal frame adapts to the Minkowski setting to give the columns of P. Using $(\xi\eta)^2 \leq \xi^2(\eta^2 + \zeta^2)$, we find

$$(\alpha\beta - \gamma\delta)^2 \leq (\alpha^2 - \gamma^2 - 1)(\beta^2 - \delta^2 + 1)$$

or after rearranging terms

$$(\alpha\delta - \beta\gamma)^2 \leq \alpha^2 - \gamma^2 - \beta^2 + \delta^2 - 1. \tag{2.8}$$

But this is the condition for ζ to exist given that ξ exists (from (2.4)) and η is determined by (2.7). If these conditions are satisfied then P exists and is unique (after the rotations to make $v_4 = v_5 = w_5 = 0$) up to choice of the remaining columns, which do not change e.

We have proved that Q is the top left corner of a future-preserving isometry iff (2.4), (2.8) and $\alpha > 0$, and further if these conditions hold then Q determines e up to isometry fixing r as claimed.

There is one other consequence that we need to use, namely

$$\alpha^2 - \beta^2 \geq 1. \tag{2.9}$$

This is most easily proved by appealing to symmetry. The top row of Q must also extend to a time-like unit vector. P is an isometry matrix in the Minkowski pseudo-metric iff $P^T M P = M$ where M is the diagonal matrix with entries $(-1, 1, 1, 1)$. Inverting and multiplying by P on the left and P^T on the right we find $M = PMP^T$ since $M^{-1} = M$; i.e. P^T is also an isometry matrix.

From (2.9) using $\alpha > 0$ we deduce $\alpha > |\beta|$.

2.3 Causal null geodesics

The coefficients A, B, C, D in (2.3) are all non-negative. To see this, add (2.5) and (2.6) and twice (2.7) to obtain

$$-(\alpha+\beta)^2+(\gamma+\delta)^2+(\xi+\eta)^2+\zeta^2=0,$$

and use $\alpha > |\beta|$ to deduce that $\alpha+\beta \geq |\gamma+\delta|$, and so $A, C \geq 0$. Similarly, (2.5) plus (2.6) minus twice (2.7) gives

$$-(\alpha-\beta)^2+(\gamma-\delta)^2+(\xi-\eta)^2+\zeta^2=0,$$

so using $\alpha > |\beta|$ again, $\alpha-\beta \geq |\gamma-\delta|$ and thus $B, D \geq 0$. It is worth noting that if either of $B, C = 0$ then $AD = 1$ so $A, D > 0$, because (2.8) can be written

$$(AD-BC)^2 \leq 2(AD+BC)-1 \tag{2.10}$$

or $$(AD-1)^2 \leq BC(2+2ABCD+BC).$$

Similarly, if either of $A, D = 0$ then $BC = 1$ so $B, C > 0$.

Note that (2.10) can be written in the form $(a^2-1)(1-d^2) \geq 0$ with $a = \sqrt{AD}+\sqrt{BC}$, $d = \sqrt{BC}-\sqrt{AD}$ which transforms the constraints on A, B, C, D to $a \geq 1$, $|d| \leq 1$, a form we shall use later. (We hope no confusion results from the temporary reuse of the symbol a since we have set the de Sitter radius to 1.)

Given U, the solutions of (2.3) can be written

$$T = \frac{U \pm \sqrt{BD+(1-BC-AD)U^2+ACU^4}}{B-AU^2} \tag{2.11}$$

(or just $T = 2(CU-D/U)$ if $B = AU^2$). The square root is always real. To see this, use $AC \geq 0$. If $AC = 0$ then one of $A, C = 0$ and in either case using the remark of the previous paragraph the argument of the square root is $BD \geq 0$. If $AC > 0$ then minimising over U^2, the argument of the square root is at least

$$\frac{2(BC+AD)-1-(AD-BC)^2}{4AC} \geq 0$$

by (2.10).

We are interested in the causal connecting null geodesics, i.e. those for which

$$\sinh(t) \geq \alpha \sinh(u)+\beta \cosh(u).$$

This condition can be expressed as

$$T-1/T \geq (AU-B/U)+(CU-D/U).$$

Use (2.3) to write $CU-D/U = 2T+(AU-B/U)T^2$ and substitute this into the causality condition to obtain

$$1/T \leq B/U-AU.$$

Since $U, T > 0$, this requires $B > AU^2$ and then the causality condition is

$$T \geq U/(B-AU^2).$$

This selects the positive square root in (2.11) and automatically gives $T > 0$.

To summarise, there is a causal connecting geodesic from given u on the emitter iff $e^u = U < \sqrt{B/A}$ (interpreted as $+\infty$ if $A = 0$); its arrival time t is given by

$$e^t = T = \frac{U+\sqrt{BD+(1-BC-AD)U^2+ACU^4}}{B-AU^2}.$$

A typical Mathematica plot of the resulting relation between u and t is reproduced in Figure 1.

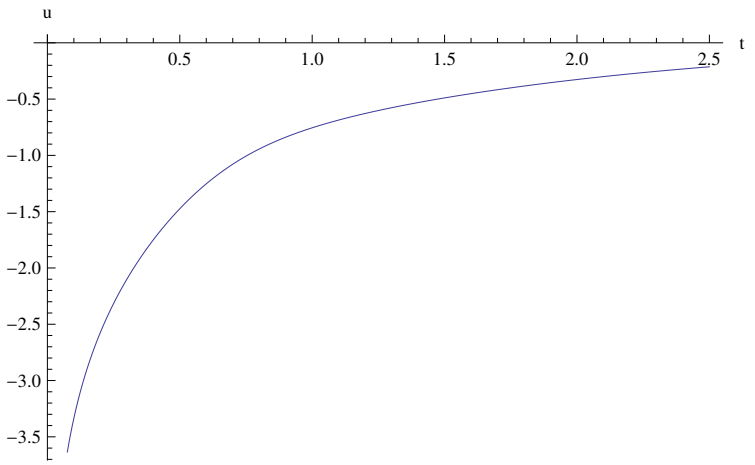

Figure 1. Emitter time u against receiver time t (with origins shifted to u^*, t^*)

Note that as $u \to -\infty$ ($U \to 0$) then $T \to T^* := \sqrt{D/B}$, so $t \to t^* = \frac{1}{2}\log(D/B)$. So if $D > 0$ there is a first receiver time t^* at which the emitter becomes visible. For U small (u large negative), $T \sim (\sqrt{BD} + U)/B$, so

$$t \sim t^* + U/\sqrt{BD} = t^* + e^u/\sqrt{BD}. \tag{2.12}$$

Similarly, $t \to +\infty$ as $U \to \sqrt{B/A}$, i.e. as $u \to u^* = \frac{1}{2}\log(B/A)$. So there is a last emitter time u^* which can be seen by the receiver. There is such a family of causal connecting null geodesics iff $B > 0$. In Figure 1, the origins of t and u have been shifted to t^*, u^* respectively.

The figure suggests that t is a monotonically increasing function of u where defined (i.e. for $u < u^*$) and this can be proved as follows. There is a similarly derived formula for U in terms of T, namely

$$U = \frac{BT^2 - D}{T + \sqrt{CD + (1 - AD - BC)T^2 + ABT^4}},$$

which is the inverse function to that defined by (2.11). This is what we actually used to plot Figure 1 and is reproduced in the Mathematica script in section 3. Therefore both functions are diffeomorphisms on the relevant open intervals and therefore monotonic.

2.4 Pictures

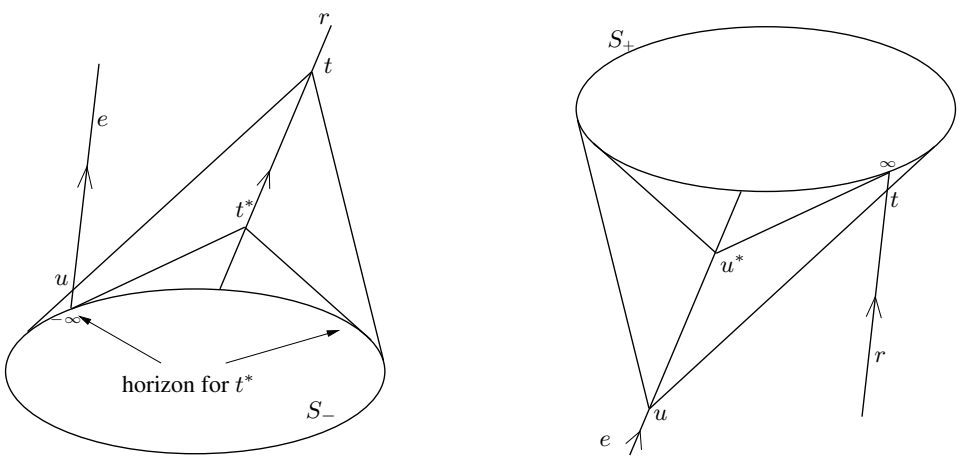

Figure 2. Left: first encounter Right: last contact

The pictures in Figure 2 illustrate these results. They are projective pictures in which the light spheres, which are at infinity in deS, are represented by finite spheres, and geodesics are represented by straight lines with null geodesics tangent to the light spheres. One can think of these pictures as obtained by projecting from the origin in Minkowski space onto a hyperplane not through the origin. The left picture gives the geometric meaning of t^* namely the point on r where the past light cone passes through the point of intersection of e with S_-, the light sphere at time $-\infty$. At this point e comes over the "horizon" of r, i.e. the sphere of contact of the past light cone with S_-. In a very short time interval of r light arrives from the past of e back to $u = -\infty$ causing the gamma-ray burst. The right picture gives the dual interpretation of u^*. This is the point where the forward light cone from e passes through the point of intersection of r with S_+, the light sphere at time $+\infty$, and after which no light passes from e to r. A small interval in e just before u^* corresponds to a time interval stretching to $t = +\infty$ in r, an infinite redshift. The pictures make it obvious that the correspondence $u \to t$ is monotonic.

2.5 Redshift and flux

The redshift z is given by

$$1 + z = \frac{dt}{du} = \frac{U}{T}\frac{dT}{dU}, \tag{2.13}$$

which can be written as a function of T, U by differentiating (2.3)

$$1 + z = \frac{AUT + BT/U - CU/T - D/(TU)}{-AUT + BT/U - CU/T + D/(TU)}. \tag{2.14}$$

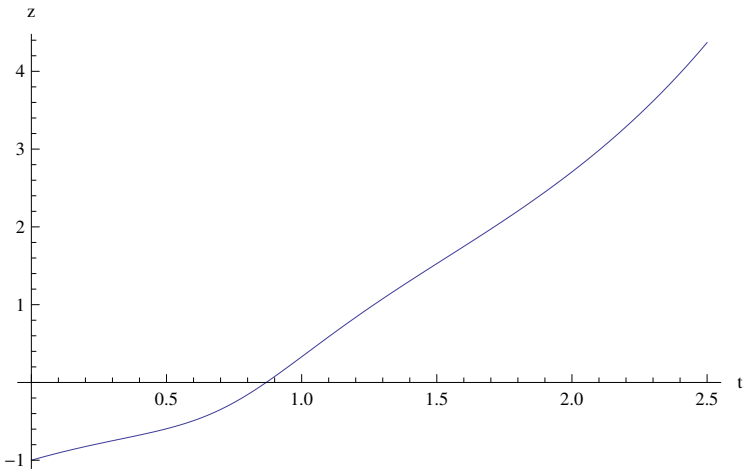

Figure 3. Redshift z (blueshift when negative) against receiver time t

A typical graph is sketched in Figure 3. As $u \to -\infty$, $1 + z \to 0$, so the first light received at t^* is infinitely blueshifted. For large negative u, $1 + z \sim U/\sqrt{BD} \sim t - t^*$. So the ratio of received to emitted frequency decreases asymptotically like $1/(t - t^*)$ as t leaves t^*. Similarly as $t \to \infty$, $z \to \infty$ like $\sqrt{AB}e^t$ and the last light received is infinitely redshifted.

The received flux Φ per unit receiver time and unit perpendicular area, given emitter power P per unit emitter time in the relevant solid angle, can be calculated by the formula (e.g. [Per])

$$\Phi = \frac{P}{((1 + z)\rho)^2}, \tag{2.15}$$

where ρ is the "corrected luminosity distance", which takes account of the geometric spreading of rays ($(1 + z)\rho$ is known as the luminosity distance). The emitter power P is, in general, a function of emitter time u.

The focussing equation [MTW, page 582, eq 22.37], [Per] for the cross-sectional area $\mathcal{A}$ of a

bundle of null geodesics with no rotation (since the bundle started with none) is:

$$\frac{d^2\mathcal{A}^{1/2}}{d\lambda^2} = -\left(|\sigma|^2 + \tfrac{1}{2}R_{ij}k^ik^j\right)\mathcal{A}^{1/2}, \tag{2.16}$$

where λ is an affine parameter, $|\sigma|$ is shear, R_{ij} is the Ricci tensor and k is a vector along the bundle. But Einstein's equation for a vacuum is $R_{ij} = \Lambda g_{ij}$ where g_{ij} is the metric and hence the second term in the bracket on the RHS reduces to Λ times the length of k which is zero since k is null. Further, in de Sitter space, the Weyl C tensor is zero and hence from the Sachs–Newman–Penrose equation [NP] for the evolution of shear

$$\frac{d\sigma}{d\lambda} = 2\tilde{\rho}\sigma + \psi_0 \quad \text{where} \quad \tilde{\rho} = -\frac{d}{d\lambda}\log\mathcal{A}^{1/2} \quad \text{and} \quad \psi_0 = C_{abcd}k^ak^c,$$

the shear of a bundle of rays remains zero if it starts zero, which is the case for a bundle originating from a point. Therefore $\mathcal{A}^{1/2}$ is an affine function of λ and we deduce that the luminosity distance is the change in affine parameter along the null geodesic, scaled to the change in emitter time at the emitter.

Thus we have

$$\rho = (\lambda_r - \lambda_e)\Big/\frac{\partial\lambda}{\partial u}$$

for any affine parameter λ, where λ_r denotes the value of λ when the null geodesic hits the receiver and λ_e the value when it leaves the emitter, and u is time in the local emitter frame. Now we can take λ to be any linear function of coordinates $(x_0, \ldots, x_4)$ whose level sets are transverse to the null geodesic, because null geodesics in deS are null geodesics in M^5 where any transverse linear function is an affine parameter. The simplest formula is obtained by applying P^{-1} to put the emitter in the standard position $e_0 = \sinh(u), e_1 = \cosh(u)$ and the other coordinates zero, and then taking $\lambda = x_0 - x_1$, because then we find $\lambda_e = -e^{-u} = -1/U$ on the emitter, and the denominator is $1/U$.

Now $P^{-1} = M^{-1}P^TM$, where $M = \mathrm{diag}(-1, 1, 1, 1, 1)$ and the top left corner of P^{-1} is $\begin{pmatrix} \alpha & -\gamma \\ -\beta & \delta \end{pmatrix}$. Thus λ on the receiver is $\alpha\sinh(t) - \gamma\cosh(t) + \beta\sinh(t) - \delta\cosh(t) = AT - C/T$ and we have

$$\rho = (AT - C/T + 1/U)U = 1 + (AT - C/T)U. \tag{2.17}$$

The equation makes it clear that if $B, D > 0$ then ρ starts at 1 (when $U = 0$, equivalently $u = -\infty$) and goes to infinity as $u \to u^*, t \to \infty$. ρ decreases initially if $BC > AD$ whereas it goes to infinity monotonically if $AD \geq BC$. It may seem strange that ρ begins from 1 rather than ∞, but isotropic emission in the emitter frame maps to emission strongly beamed along the emitter forward velocity in the receiver frame. As $t \to +\infty$, ρ is asymptotically proportional to $T = e^t$. So

$$\Phi \sim \begin{cases} \dfrac{P}{(t-t^*)^2} & \text{for } t \text{ just after } t^* \\[2ex] \dfrac{Pe^{-4t}}{A^2B^2} & \text{as} \quad t \to +\infty. \end{cases} \tag{2.18}$$

Typical plots of Φ against t (intensity plot) and z against ρ (Hubble plot) are given in Figure 4.

2.6 Blueshift period

We now look for points t_0 where $z = 0$ (i.e. $dt/du = 1$), which represent the boundary where blueshift changes to redshift or vice versa. Differentiate (2.3) with respect to t and set $du/dt = 1$ (or $z = 0$ in (2.14)) then the B and C terms both disappear and we obtain:

$$-AUT + D/UT = 0 \qquad \text{or} \qquad UT = \sqrt{D/A}. \tag{2.19}$$

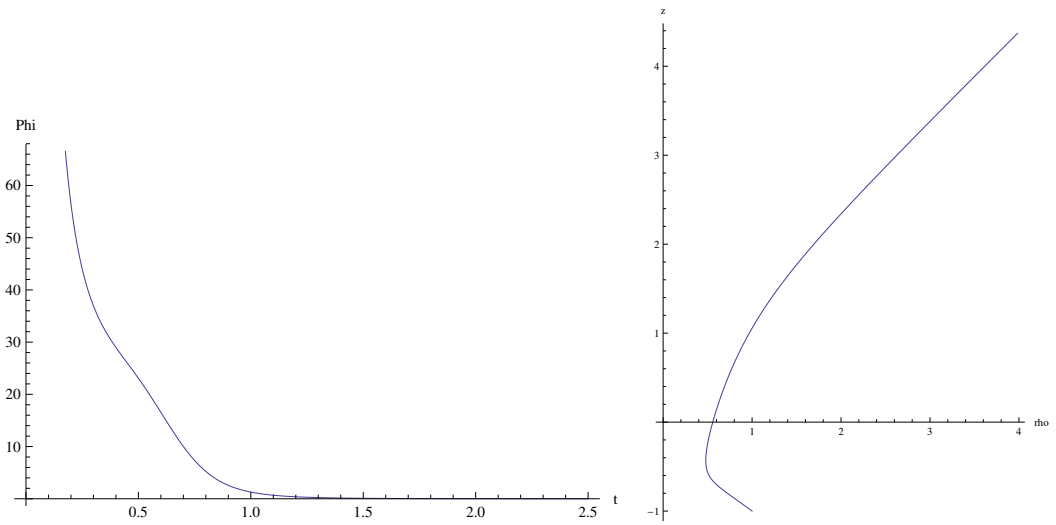

Figure 4. Left: received flux `Phi` as a function of receiver time t; right: Hubble plot (redshift z against corrected luminosity distance `rho`).

Substituting back in (2.3) yields:

$$-\sqrt{DA} + B\sqrt{A/D}T^2 + C\sqrt{D/A}T^{-2} - \sqrt{AD} = 2,$$

or

$$B\sqrt{A/D}T^4 - 2(\sqrt{AD}+1)T^2 + C\sqrt{D/A} = 0, \tag{2.20}$$

which is a quadratic in T^2. The larger solution T_0 is on the causal light ray. So

$$T_0^2 = \frac{1+\sqrt{AD}+\sqrt{1+2\sqrt{AD}+AD-BC}}{B\sqrt{A/D}},$$

where $T_0 = e^{t_0}$.

Thus we can see that the received light starts at t^* infinitely blueshifted and drops to zero blueshift at t_0 and thereafter is redshifted. The difference $t_B = t_0 - t^*$ is the *blueshift period* given by:

$$t_B = \frac{1}{2}\log\frac{1+\sqrt{AD}+\sqrt{1+2\sqrt{AD}+AD-BC}}{\sqrt{AD}}. \tag{2.21}$$

This can be rearranged as

$$t_B = \log\left(\sqrt{\frac{1+\sqrt{AD}+\sqrt{BC}}{2\sqrt{AD}}} + \sqrt{\frac{1+\sqrt{AD}-\sqrt{BC}}{2\sqrt{AD}}}\right),$$

a form that we shall see again later.

2.7 Case analysis

There is apparently a 4–parameter family of pairs of time-like geodesics to consider, parametrised by $(\alpha,\beta,\gamma,\delta)$ or equivalently (A,B,C,D), but one can shift the origins of u and t along the geodesics so really there is only a two-parameter family of pairs, the two important parameters being the products AD and BC.

This can also be seen geometrically by using the equivalence with hyperbolic 4–space. Recall that a geodesic in deS is the intersection of the hyperboloid with a plane through the origin. The plane determining a time-like geodesic meets the copy $\mathbb{H}^4$ of hyperbolic 4–space, given by $x_0^2 - \sum_{i=1}^4 x_i^2 = 1$ and $x_0 > 0$, in a (hyperbolic) line. The group of isometries of deS (as a

Lorentzian manifold) is precisely the same as the group of hyperbolic isometries of $\mathbb{H}^4$. So we can use hyperbolic arguments to analyse the various cases that arise. Now two lines in hyperbolic space have a unique common perpendicular corresponding to the points of closest approach except in the *intersecting* case (when they have a point in common), which we can think of as being the common perpendicular but of zero length, and in the *asymptotic* case when they meet on the sphere at infinity. Up to hyperbolic isometry there is only one asymptotic case (think of the upper half-space model and two parallel lines perpendicular to the boundary: there are isometries which move the boundary by an arbitrary similarity transformation). The other cases are characterised (again up to hyperbolic isometry) by (a) the length of the common perpendicular and (b) the angle between the two lines after translation along the common perpendicular to be intersecting.

There are three *degenerate* cases, (A) asymptotic, (B) intersecting and (C) parallel (angle zero). These are the cases in which the planes in M^5 determining the lines span a 3–space (rather than a 4–space, as they would in general position) and hence have a line through the origin in common. The cases above correspond to this line (A) lying in the light cone (B) being time-like (meeting $\mathbb{H}^4$ in a point) and (C) being space-like (meeting deS in two antipodally opposite points).

Going back to thinking about deS rather than $\mathbb{H}^4$, a little care is needed because a plane through the origin meets deS in a pair of antipodally opposite geodesics. Case (A) is still naturally thought of as the asymptotic case, though now there are four subcases: future and past asymptotic (meeting on the future resp. past light cone) and the antipode versions where the lines are future or past asymptotic to opposite points of the light cone. Case (B) is now naturally the "parallel" case: take the point of intersection to be the centre of $\mathbb{H}^4$ and then the geodesics are both vertical in the hyperboloid. Finally case (C) is now the intersecting case.

In terms of the conditions on Q found in section 2.2, degeneracy is the case when $\zeta = 0$, i.e. when $(\alpha^2 - \gamma^2 - 1)(\beta^2 - \delta^2 + 1) = (\alpha\beta - \gamma\delta)^2$. It is easy to characterise the various cases: Define the discriminant $\Delta = \frac{\alpha\eta - \beta\xi}{\gamma\eta - \delta\xi}$ where $\xi^2 = \alpha^2 - \gamma^2 - 1$, $\eta^2 = \beta^2 - \delta^2 + 1$ (since $\zeta = 0$). Then if $|\Delta| > 1$ the intersection is virtual (in $\mathbb{H}^4$), if $|\Delta| = 1$ then the geodesics are asymptotic, and if $|\Delta| < 1$ then the intersection is real. The proof is to look for the common line in the $(0,1)$–plane and its image under P (spanned by the vectors $(\alpha, \gamma, \xi, 0, 0)$ and $(\beta, \delta, \eta, 0, 0)$) and think projectively. Note that the case $\Delta = \infty$ is the case of parallel geodesics as described above.

2.8 A parametrisation of nearly all cases

We now give an explicit 2–parameter family of matrices Q which cover all cases (as usual up to isometry) except the asymptotic cases which need to be analysed separately. Start with the standard geodesic r, and to define e, first rotate by an angle θ in the $(1,2)$–plane and second shear (hyperbolic rotation by ϕ, formula below) in the $(0,3)$–plane. These motions commute and thinking hyperbolically you are taking a line, rotating it in a containing plane and then moving it away and by symmetry the motion is along a line perpendicular to the rotation plane. Thus this gives the general (nondegenerate) pair.

In terms of matrices, the motion is given by

$$P = \begin{pmatrix} a & 0 & 0 & b & 0 \\ 0 & c & -s & 0 & 0 \\ 0 & s & c & 0 & 0 \\ b & 0 & 0 & a & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where $c = \cos(\theta)$, $s = \sin(\theta)$, $a = \cosh(\phi)$, $b = \sinh(\phi)$; c or θ parametrizes the rotation and a or ϕ parametrises the shear (hyperbolic rotation). (Here again we hope the temporary reuse of a does not cause confusion.) In terms of our standard notation the top left (2×2)–matrix is $\begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$, where $\alpha = \cosh(\phi)$ and $\delta = \cos(\theta)$. One nice feature of this parametrisation is that it takes care of the ambiguity in lifting to deS because a rotation through π which corresponds

to changing the sign of δ also corresponds to replacing the image line (emitter) in deS by the antipodally opposite line.

The analysis made in the general case simplifies enormously. The conditions on Q reduce to $\alpha \geq 1$ and $|\delta| \leq 1$ with degeneracy given by $\alpha = 1$ or $|\delta| = 1$. If $\alpha = 1$ there is no shear and the lines are intersecting in $\mathbb{H}^4$, alias "parallel" in deS. If $\delta = 1$ there is no rotation and the lines are parallel in $\mathbb{H}^4$, alias intersecting in deS, and if $\delta = -1$ they are still parallel in $\mathbb{H}^4$ but pass through antipodally opposite points in deS. The null geodesic condition reduces to:

$$-\alpha \sinh(u)\sinh(t) + \delta \cosh(u)\cosh(t) = 1. \tag{2.22}$$

We can read $2A = 2D = \alpha - \delta$ and $2B = 2C = \alpha + \delta$. Thus the condition $B = 0$ for the non-existence of causal connecting geodesics is $\alpha + \delta = 0$ which only happens if $\alpha = 1$ and $\delta = -1$. In this case the lines are parallel and antipodally opposite in deS and it is easy to see geometrically that there are no connecting null geodesics causal or otherwise. Later we shall find just one other case (an asymptotic case not covered by the (α, δ) parametrisation) where there are no connecting causal null geodesics.

It is easy to find t_0 explicitly (where $z = 0$): Differentiate (2.22) and use $du/dt = 1$ to get

$$(\delta - \alpha)(\cosh(u)\sinh(t) + \sinh(u)\cosh(t)) = 0,$$

which implies $\tanh(u) = -\tanh(t)$. Thus $u = -t$ where $z = 0$, which can also be seen by the symmetry of (2.22). Now substitute in (2.22) to get $\alpha \sinh^2(t_0) + \delta \cosh^2(t_0) = 1$, which gives:

$$(\alpha + \delta)\cosh^2(t_0) = \alpha + 1 \qquad \text{or} \qquad t_0 = \operatorname{arccosh}\left(\sqrt{\frac{\alpha+1}{\alpha+\delta}}\right) \tag{2.23}$$

Clearly the causal value is the positive value. Note that $t_0 \to 0$ as $\alpha \to \infty$.

We already know that $t^* = \frac{1}{2}\log((\alpha-\delta)/(\alpha+\delta))$ (since $t^* = \frac{1}{2}\log(D/B)$, subsection 2.3) but we can recover this directly as follows. Set $u = -\infty$ in (2.22) to get $\alpha \sinh(t^*) + \delta \cosh(t^*) = 0$ and hence:

$$\tanh(t^*) = -\delta/\alpha \qquad \text{or} \qquad t^* = \operatorname{arctanh}(-\delta/\alpha), \tag{2.24}$$

and the previous formula follows from the standard formula $\operatorname{arctanh}(x) = \frac{1}{2}\log((1+x)/(1-x))$. Note that $t^* \to 0$ as $\alpha \to \infty$ as well and we deduce that the blueshift period $t_B = t_0 - t^*$ also $\to 0$ as $\alpha \to \infty$. This can also be deduced directly from (2.22) because when α is large we must have $\sinh(u)\sinh(t)$ small.

There is also a standard formula for $\operatorname{arccosh}(x) = \log(x + \sqrt{x^2-1})$ so that we get a formula for t_B in the general case:

$$\begin{aligned} t_B &= t_0 - t^* \\ &= \log\left(\sqrt{(\alpha+1)/(\alpha+\delta)} + \sqrt{(1-\delta)/(\alpha+\delta)}\right) - \log\left(\sqrt{(\alpha-\delta)/(\alpha+\delta)}\right) \\ &= \log\left(\sqrt{(\alpha+1)/(\alpha-\delta)} + \sqrt{(1-\delta)/(\alpha-\delta)}\right). \end{aligned} \tag{2.25}$$

It is important to point out that the meaning of the parameters α and δ is quite different in deS than it is in $\mathbb{H}^4$. In deS it is natural to regard δ as separation of the orbits and α as relative velocity. Thus we have short blueshift period when the orbits have a very high relative velocity. In Figure 5 we have reproduced a Mathematica plot for t_B as a function of θ and ϕ. You can see that t_B takes all values between 0 and ∞.

2.9 Received flux and blueshift period

For α large and $\delta \in [-1, +1]$ (2.25) gives the expansion:

$$t_B = \sqrt{\frac{1-\delta}{\alpha}} + \frac{\delta}{\alpha} + O\left(\frac{1}{\alpha^{3/2}}\right),$$

so the shortest blueshift periods are for α large and δ near 1, though for any value of δ, $t_B \to 0$ as $\alpha \to \infty$.

[More generally (in terms of (A, B, C, D)) we have

$$t_B \sim \sqrt{\frac{1 + \sqrt{AD} - \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}}$$

for $\sqrt{AD}$, $\sqrt{BC}$ large and with bounded difference, and t_B is small if AD and BC are large, with $\sqrt{BC}$ near $1 + \sqrt{AD}$.]

From (2.18) the received flux integrated between t^* and a nearby t behaves like

$$\int_{t^*}^{t} \frac{1}{(t - t^*)^2} P(u) dt,$$

where $P(u)$ is the power at emitter time u. But from (2.12) we have $e^u = U \sim (t - t^*)\sqrt{BD}$. Changing variables and simplifying, the received energy per unit area is asymptotically given by

$$\sqrt{BD} \int_{-\infty}^{u(t)} e^{-u} P(u) du.$$

We suppose that this integral converges as for example it does if $P(u)$ is integrable and has compact support. Then the dependence on parameter is purely the factor $\sqrt{BD} = \frac{1}{2}\sqrt{\alpha^2 - \delta^2}$. Thus the short blueshift regime (α large) is correlated with high received flux.

Exact treatment, rather than the asymptotics for small $t - t^*$, yields received energy per unit area up to time t

$$\int_{-\infty}^{u(t)} \frac{DP(u)}{T\rho^3} \left(e^{-u} + \frac{T}{D} - \frac{Ce^u}{D} \right) du. \tag{2.26}$$

This makes clear that a strong enhancement is produced by ρ decreasing initially. Recall that

$$\rho = 1 + (AT - \frac{C}{T})U,$$

so it has fastest initial decrease if $BC \gg AD$ (the initial T is $T^* = \sqrt{D/B}$). This is the regime $\alpha\delta \gg 1$, but δ is limited to $|\delta| \leq 1$, so is the regime α large, δ near 1, i.e. the regime of short blueshift period.

This is a very significant result physically because it may explain why the observed gamma-ray bursts have very short blueshift periods. As the sources causing the bursts are a long distance away (meaning that the corrected luminosity distance starts at the de Sitter radius) then we will observe them only if they are concentrated by this effect. The long blueshift bursts are still there but so weak that they merely contribute to the background radiation.

There is an additional, even more significant, effect that biases the distribution of blueshift period towards 0, which we will explain in section 4.1.

2.10 The asymptotic cases

None of the asymptotic cases are important for gamma-ray bursts, but we are including them to make the analysis complete. As we saw earlier, there are exactly four asymptotic cases.

Starting with the case of geodesics passing through the same point on the bottom light cone, we can take Q to have the simplest form, namely $\begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$, which makes e pass through the (projective) point $(-1, 1, 0, 0, 0)$, i.e. be past asymptotic to r.

An aside: it is not hard to prove by calculation that up to isometry any geodesic e past asymptotic to r comes from this matrix, which avoids the appeal to hyperbolic geometry made earlier. We can assume that P has first two columns of the form $(\alpha, \gamma, \xi, 0, 0)^T$ and $(\beta, \delta, \eta, 0, 0)^T$ and that $Pv = \lambda v$ where $v = (-1, 1, 0, 0, 0)^T$. It can be checked that postmultiplying by

$A = \begin{pmatrix} c & s \\ s & c \end{pmatrix}$, where $c^2 - s^2 = 1$, which is a general reparametrization of e, has the effect of multiplying λ by $c - s$ and hence we can take $\lambda = 1$. Then using $Pv = v$, and solving the equations for orthonormality, one quickly finds that Q must have the form $\begin{pmatrix} 1+q & q \\ -q & 1-q \end{pmatrix}$ where $\xi^2 = 2q$. But again it can be checked that conjugating by A, which leaves r invariant, has the effect of replacing q by $(s-c)^2 q$ and hence we can choose q to have any positive value, e.g. 1.

We now read off: $A = 2, B = C = 1, D = 0$ and hence $t^* = \frac{1}{2}\log(D/B) = -\infty$ and the receiver sees the emitter for all t. From (2.3) we read the relationship for received time against transmitted time

$$T = \frac{U + \sqrt{2}U}{1 - 2U^2} = \frac{U}{1 - \sqrt{2}U}, \tag{2.27}$$

and then from (2.19) we see that $U = T = 0$ when $z = 0$ and hence $t_0 = -\infty$ as well and there is no blueshift period, i.e. the light is always redshifted. Differentiating (2.27) we find

$$\frac{dT}{dU} = \frac{1}{(1 - \sqrt{2}U)^2} = \frac{T^2}{U^2},$$

and hence redshift

$$1 + z = \frac{U}{T}\frac{dT}{dU} = \frac{T}{U} = e^{t-u},$$

and we have perfect exponential dependence of redshift on time difference. This is not at all surprising because e and r are time-lines in the standard expanding observer field Exp and share a coherent time frame [MR, MR2].

This is the case analysed in [We] (and reiterated in [We30]). It seems to us to be a great mistake that he introduced his coherency postulate that all visible matter must have been visible at all past times, thereby forcing himself to restrict to this case.

The case of geodesics asymptotically meeting on the future light sphere is obtained by reversing the time coordinate and we find a perfect exponential blueshift depending on time difference. Again this is a standard contracting observer field.

Turning now to the case when the two geodesics pass through antipodally opposite points on the bottom light sphere, we can take the matrix to be $\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ and then $C = 2, A = D = 1, B = 0$ and $t^* = \infty, u^* = -\infty$ so there are no null geodesics from e to r.

Reversing time, i.e. future asymptotic to opposite points on the top light sphere, we can take the matrix to be $\begin{pmatrix} 2 & -1 \\ -1 & 0 \end{pmatrix}$ and $B = 2, A = D = 1, C = 0$ and both t^* and u^* are finite: $t^* = \frac{1}{2}\log(1/2)$ and $u^* = \frac{1}{2}\log(2)$.

So we have exactly two cases where there are no causal geodesics: antipodally opposite and past asymptotic to antipodally opposite.

3 Mathematica plots

3.1 A typical case

Below is the basic notebook. It plots four relationships for choices of the two control parameters θ and ϕ defined above. The names for the variables are not quite the same as in the paper. `theta` and `phi` are θ and ϕ and similarly `rho` and `Phi` are ρ and Φ; `a` and `d` are α and δ; `A` and `B` are A and B but `CC` and `DD` are C and D (because `C` and `D` have reserved meanings in Mathematica). The four output graphs (Figures 1, 3 and 4, given in Sections 2.3 and 2.5) are for the case $\theta = 0.5$ and $\phi = 1$.

```
theta = 0.5; phi = 1;
a := Cosh[phi]; d := Cos[theta];
```

```
A := (a - d)/2; B := (a + d)/2; CC := B; DD := A;
TT := Sqrt[DD/B]; UU := Sqrt[B/A]; tmax = 2.5;
T := TT*Exp[t];
U := (B*T - DD/T)/(1 + Sqrt[1 + (A*T - CC/T)*(B*T - DD/T)]);
Plot[Log[U/UU], {t, 0, tmax}, PlotPoints -> 1000,
 AxesLabel -> {"t", "u"}, AxesOrigin -> {0, 0}]
rho := 1 + (A*T - CC/T)*U;
z := 2*(A*T - DD/(T*U^2))/((B*T + DD/T)/U^2 - (A*T + CC/T));
ParametricPlot[{rho, z}, {t, 0, tmax}, PlotPoints -> 1000,
 AxesLabel -> {"rho", "z"}, AxesOrigin -> {0, 0}]
Phi := ((1 + z)*rho)^{-2};
Plot[Phi, {t, 0, tmax}, AxesLabel -> {"t", "Phi"}, PlotPoints -> 1000]
Plot[z, {t, 0, tmax}, AxesLabel -> {"t", "z"}, PlotPoints -> 1000]
```

3.2 Blueshift period

Figure 5 is a 3D plot of the blueshift period against the parameters θ, ϕ. The code is:

```
a := Cosh[phi]; d := Cos[theta];
Plot3D[Log[Sqrt[(a + 1)/(a - d)] + Sqrt[(1 - d)/(a - d)]],
 {phi, 0,  5}, {theta, 0, Pi}, AxesLabel -> {"", ""},
 PlotRange -> {0, 2}]
```

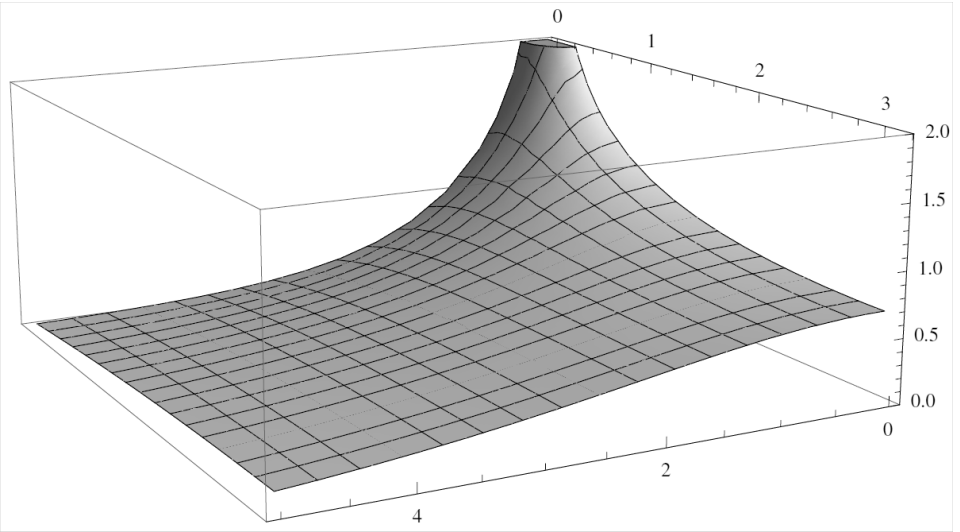

Figure 5. Blueshift period (vertical) as a function of θ (top axis) and ϕ (bottom axis)

3.3 Intensity plots

We now concentrate on intensity plots with ϕ large, which therefore have short blueshift period. In each of the plots in Figure 6 we have $\phi = 4$ and we have reduced `tmax` to 0.15. A noticeable feature is the double-hump for small enough θ.

3.4 Observations

In Figure 7 we have reproduced the selection of observations from Wikipedia (a similar set appears in Figure 1 of [Mes]). Apart from modulation, the theoretical curves in Figure 6 are a good fit for most of these observations. By varying the two parameters θ and ϕ we can reproduce the typical single peak (possibly expanded) and double peak observed gamma-ray bursts.

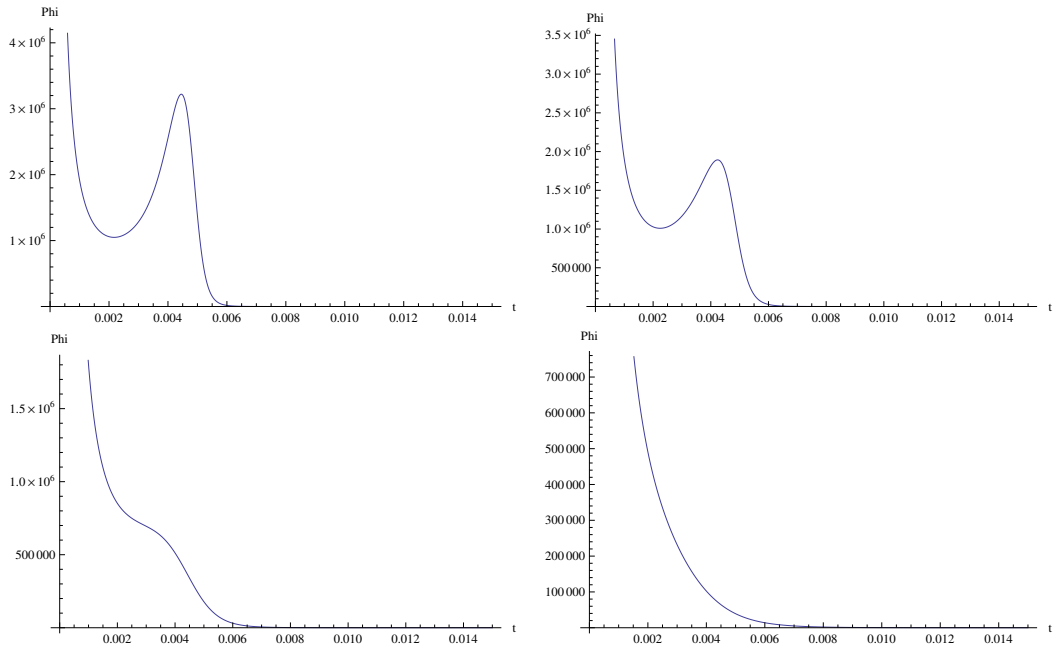

Figure 6. Received flux `Phi` as a function of receiver time `t`, for $\phi = 4$ and $\theta = 0.15, 0.3, 0.4, 0.8$ (top-left, top-right, bottom-left and bottom-right respectively)

The spiky modulation observed in some cases of GRB could be the result of non-trivial variation in the emitter power $P(u)$ over long periods of emitter time, which are compressed into a short period of receiver time. Indeed, it would be interesting to seek to infer ϕ, θ and the function P from observed light-curves and to see whether the inferred power P fits with believed models for the life course of potentially relevant emitters.

A second contribution to the spiky modulation can come from lensing effects for deviations from de Sitter space, as will be described in Section 5.1.

4 Further considerations

4.1 Distribution of parameters

What distribution of the parameters θ and ϕ (equivalently $a = \cosh\phi$ and $c = \cos\theta$) should we expect to see? If emitters are distributed uniformly in deS in positions and velocities then we will deduce the distribution

$$d\mu = \sinh^2\phi \sin\theta \, d\phi \, d\theta. \tag{4.1}$$

The idea is that emitter geodesics are given by applying an isometry M of deS (without loss of generality, future and orientation-preserving) to any time-like reference geodesic, for example the receiver geodesic. The isometries M form a group $SO^+(1,4)$. Up to scaling, there is a unique measure invariant under the action (left or right) of the group, called Haar measure. This is the natural measure to use. It is non-normalisable because the group is non-compact, but that does not stop us talking about uniform distribution; compare the problem of deciding what we mean by points uniformly distributed in the Euclidean plane.

Rather than time-like geodesics in deS, we find it easier to think about the corresponding geodesics in $\mathbb{H}^4$. As before, we take $\mathbb{H}^4$ to be the upper hyperboloid $-x_0^2 + \sum_{j=1}^4 x_j^2 = -1$, $x_0 > 0$, in Minkowski 5-space. Time-like geodesics in deS (strictly speaking, antipodal pairs of them) and all geodesics in $\mathbb{H}^4$ are defined by the intersections of 2-planes through 0 of slope greater than 45°. We associate one to the other if they are defined by the same 2-plane. We take our reference geodesic r to be defined by the plane with $x_j = 0$ for $j = 2, 3, 4$.

Given two geodesics r and e in $\mathbb{H}^4$, if they are not asymptotic then there is a shortest distance between them, which we denote by ϕ because it corresponds precisely to our parameter ϕ. The

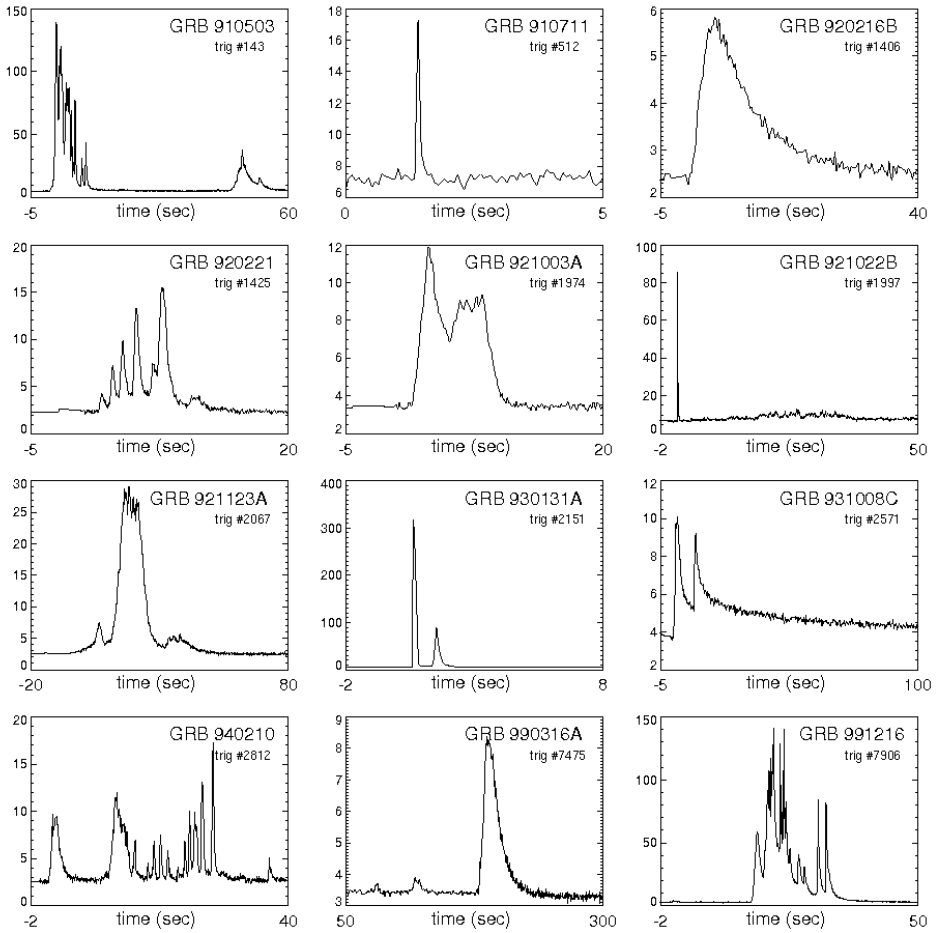

Figure 7. Gamma-ray burst observations reproduced from Wikipedia
`https://en.wikipedia.org/wiki/Gamma-ray_burst`
`GRB_BATSE_12lightcurves.png` in public domain, created by Daniel Perley

distribution for ϕ is (proportional to) $\sinh^2\phi\ d\phi$ because in $\mathbb{H}^4$ the set of points at distance ϕ from a given one on r and perpendicular to r is a 2-sphere of area $4\pi\sinh^2\phi$. The parameter θ is the angle by which e is rotated relative to r about the common perpendicular. The set of possible rotations about this perpendicular is a 2-sphere and $\theta\in[0,\pi]$ is the angle relative to the direction of r. Thus the distribution of θ is $2\pi\sin\theta\ d\theta$. The two factors are independent, hence $d\mu$ has the form claimed in (4.1).

In particular, we can predict the distribution of blue-shift periods from $d\mu$. Recall that

$$t_B=\log(\sqrt{\alpha+1}+\sqrt{1-\delta})-\log\sqrt{\alpha-\delta}, \tag{4.2}$$

where $\alpha=\cosh\phi$, $\delta=\cos\theta$. In (α,δ) variables, the natural measure is $d\mu=\sqrt{\alpha^2-1}\ d\alpha\ d\delta$. Change variable from α to $\tau=t_B$. Then $d\alpha\ d\delta=d\tau\ d\delta/J$ with

$$J=\left|\frac{\partial\tau}{\partial\alpha}\right|=\frac{1+\delta+\sqrt{1-\delta}\sqrt{\alpha+1}}{2(\sqrt{\alpha+1}+\sqrt{1-\delta})\sqrt{\alpha+1}(\alpha-\delta)}.$$

So the natural measure is

$$d\mu=\frac{2\sqrt{\alpha^2-1}(\sqrt{\alpha+1}+\sqrt{1-\delta})\sqrt{\alpha+1}(\alpha-\delta)}{1+\delta+\sqrt{1-\delta}\sqrt{\alpha+1}}d\tau\ d\delta \tag{4.3}$$

Given τ and δ we can solve (4.2) for α:

$$\alpha = \delta + \frac{2}{(e^{2\tau}-1)^2}\left(e^{2\tau} - \delta + \sqrt{e^{4\tau}(1-\delta^2) - 2e^{2\tau}\delta(1-\delta)}\right) \tag{4.4}$$

Inserting this into (4.3) we can integrate over δ to obtain the marginal distribution for τ. The formulae are too messy, however, so we content ourselves with computing the leading asymptotics for small τ. Then except for δ near 1, we have $\alpha \sim (1-\delta)/\tau^2$ and

$$d\mu \sim \frac{2(1-\delta)^2}{\tau^5} d\tau\, d\delta. \tag{4.5}$$

The region δ near 1 makes a negligible correction, so integrating over δ, the marginal distribution on τ is

$$\frac{16}{3\tau^5} d\tau. \tag{4.6}$$

This compares favourably with the distribution of observed GRB durations for duration larger than 40 seconds, but the observed distribution peaks at around 20 seconds and falls off for smaller duration [K+]. Thus we would need to invoke some cutoff in the distribution of emitters to explain the observations (compare Planck's hypothesis to correct Rayleigh-Jeans law).

4.2 Double-peak regime

Figure 6 suggests that the light curve is double-peaked for θ small enough. Indeed we prove that the light-curve is double peaked iff $\delta > \sqrt{8}/3$, i.e. $\theta < 0.34$ radians approximately (19.5°).

We obtained this condition by computing the condition for critical points of the received flux Φ with P constant, equivalently for $Q = (1+z)\rho$. Here is the working.

We wish to solve the null geodesic condition (2.22)

$$\delta \cosh u \cosh t - \alpha \sinh u \sinh t = 1$$

and the condition $Q' = 0$, where prime denotes differentiation with respect to some smooth parametrisation of the family of null geodesics. Differentiating the null condition yields

$$1 + z = \frac{dt}{du} = \frac{\alpha \sinh t \cosh u - \delta \sinh u \cosh t}{\delta \sinh t \cosh u - \alpha \sinh u \cosh t},$$

thus we choose the parametrisation so that

$$t' = \alpha \sinh t \cosh u - \delta \sinh u \cosh t \tag{4.7}$$

$$u' = \delta \sinh t \cosh u - \alpha \sinh u \cosh t. \tag{4.8}$$

Then $1 + z = t'/u'$. Also inserting the case of our two-parameter family into the formula (2.17) for ρ we obtain $\rho = t'$.

Thus $Q = t'^2/u'$ and so $Q' = 0$ is equivalent to $2t''u' = u''t'$.

Now

$$\begin{aligned} t'' &= (\alpha \cosh t \cosh u - \delta \sinh u \sinh t)t' + (\alpha \sinh t \sinh u - \delta \cosh u \cosh t)u' \\ &= (\alpha^2 - \delta^2) \sinh t \cosh t \end{aligned}$$

after expanding out and using $\cosh^2 - \sinh^2 = 1$. Similarly

$$u'' = (\alpha^2 - \delta^2) \sinh u \cosh u.$$

So $Q' = 0$ can be written as

$$\delta \cosh t \cosh u\, (2\sinh^2 t + \sinh^2 u) = \alpha \sinh t \sinh u\, (2\cosh^2 t + \cosh^2 u). \tag{4.9}$$

But $2\sinh^2 t + \sinh^2 u = 2\cosh^2 t + \cosh^2 u - 3$, so $Q' = 0$ can be rewritten as

$$(\delta \cosh t \cosh u - \alpha \sinh t \sinh u)(2\cosh^2 t + \cosh^2 u) = 3\delta \cosh t \cosh u.$$

The first parenthesis is 1 by the null condition, so

$$2\cosh^2 t - 3\delta \cosh u \cosh t + \cosh^2 u = 0.$$

Thus

$$\cosh t = \frac{1}{4}(3\delta \pm \sqrt{9\delta^2 - 8}) \cosh u. \tag{4.10}$$

There are solutions iff $\delta^2 \geq 8/9$ and $\delta > 0$ (because $\cosh \geq 1$), i.e. iff $\delta \geq \sqrt{8}/3$. Write the solutions of (4.10) as $\cosh t = D \cosh u$. We have to check that they are realisable, in particular by future-pointing null geodesics.

An independent combination of $Q' = 0$ with the null condition is obtained by putting $2\cosh^2 t + \cosh^2 u = 3 + 2\sinh^2 t + \sinh^2 u$ into (4.9) and using the null condition to obtain

$$2\sinh^2 t - 3\alpha \sinh u \sinh t + \sinh^2 u = 0.$$

Thus $\sinh t = \frac{1}{4}(3\alpha \pm \sqrt{9\alpha^2 - 8}) \sinh u$. Now $\alpha \geq 1$, so the square root is always real, but the plus sign is impossible because it would give $\sinh t / \sinh u > 1$ whereas (4.10) shows that $\cosh t / \cosh u \leq 1$ (because $\delta \leq 1$). Write the solution using the minus sign as $\sinh t = A \sinh u$, and note that $A > 0$.

Squaring (4.10) we obtain $1 + \sinh^2 t = D^2(1 + \sinh^2 u)$. Put $\sinh u = A^{-1} \sinh t$ to obtain

$$A^2(1 + \sinh^2 t) = D^2(A^2 + \sinh^2 t).$$

So

$$\sinh^2 t = \frac{A^2(1 - D^2)}{D^2 - A^2}.$$

For either choice of D we have $D^2 \leq 1$. Also for each choice, $D^2 \geq A^2$ when $\delta \geq \sqrt{8}/3$, $\alpha \geq 1$. So this equation determines $\sinh^2 t$ for a critical point of Q.

The future-pointing condition $\alpha \sinh u < \sinh t$ selects the negative solution for $\sinh t$, because $\sinh t = A \sinh u$, so this condition is $(\alpha - A) \sinh u < 0$. But $\alpha - A = \frac{1}{4}(\alpha + \sqrt{9\alpha^2 - 8}) > 0$ for $\alpha \geq 1$.

Thus we obtain no solutions if $\delta < \sqrt{8}/3$, precisely one solution if $\delta = \sqrt{8}/3$, and precisely two if $\delta > \sqrt{8}/3$.

4.3 Spectrum

It is claimed that observations show GRBs to have highly non-thermal spectrum, e.g. Figure 8 reproduced from [Gol]. In our scenario, there is no great reason to suppose the emitter spectrum to be thermal, but it is a natural null hypothesis, so let us examine the effects.

If the emitter has thermal (i.e. black-body) spectrum with temperature Θ then the received flux has thermal spectrum with temperature $\Theta/(1 + z)$. But photon count rates are often very low, of the order of less than 100 per second, thus integration over some time-interval is essential to build up a spectrum and this is indeed exactly what observers do. During this time interval, however, in our scenario the redshift changes rapidly, thus the effect of integration is to average over thermal spectra with a range of temperatures.

The standard way to present GRB spectra (as in Fig. 8) is photons/unit area/unit time/unit energy against energy E, on a log-log plot. Figure 9(a) shows a thermal spectrum plotted this way. The photon rate is

$$\frac{2E^2}{h^3c^2} \frac{1}{e^{\beta E} - 1},$$

where $\beta = 1/\Theta$ in energy units, and we took $\Theta = 1$ for the figure. The spectrum of the received flux is given by taking $\beta = (1 + z)\beta_e$ (where $\beta_e = 1/\Theta_e$ for the emitter) and multiplying by a factor $(1 + z)^2/\rho^2$ because integrated over energy the received flux is $\Phi = \frac{P}{(1+z)^2\rho^2}$ with P

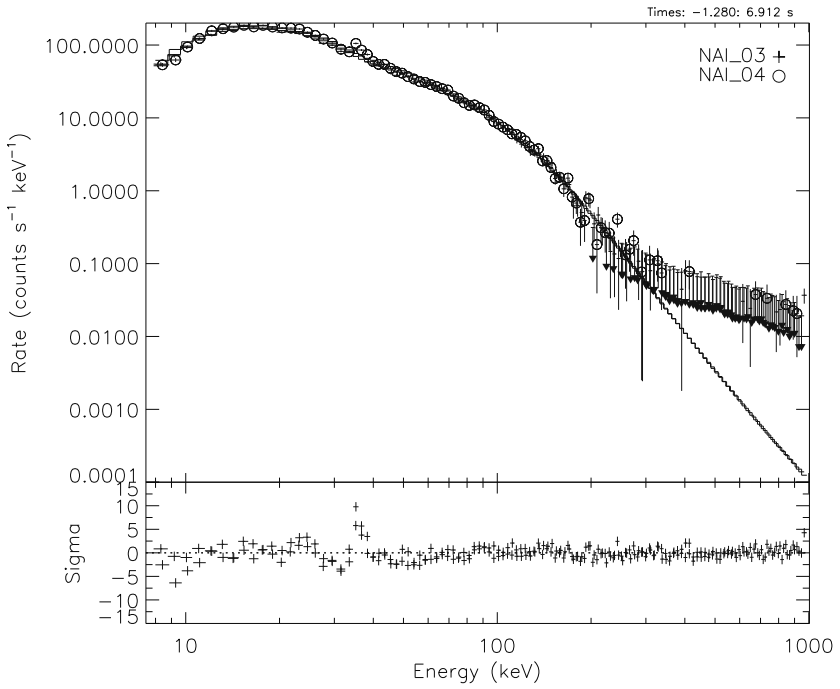

Figure 8. Spectrum of a gamma-ray burst (Figure 1 of [Gol]). ©American Astronomical Society. Reproduced with permission. The authors explain that the noisy right-hand part should be ignored.

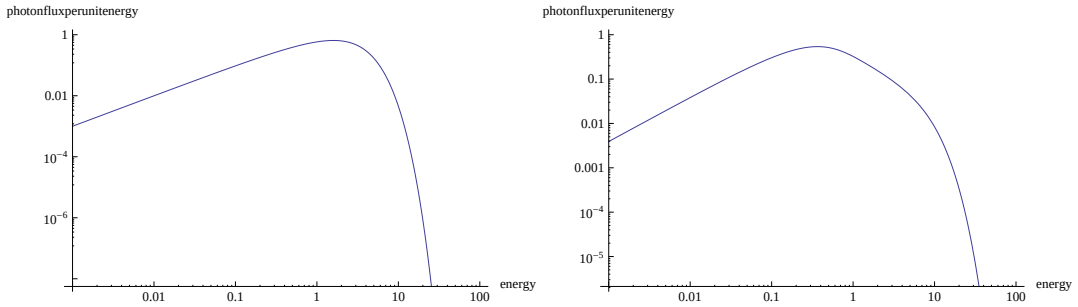

Figure 9. (a) Black body spectrum expressed in photons/unit area/unit time/unit energy against energy on a log-log plot; (b) Time-averaged received spectrum for interval $[0.375, 8.35]$.

being the emitted power/unit solid angle in the emitter frame, whereas the energy-integral of a redshifted black-body spectrum is proportional to Θ^4 (Stefan's law) so we have to compensate by a factor $(1+z)^4$. To simplify matters, we took the leading asymptotics $\rho \sim 1$ and $1+z \sim t$, for small t, being the receiver time measured from the time of first sight. Then the averaged received spectrum over time interval $[t_0, t_1]$ is

$$\frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \frac{E^2 t^2 \, dt}{e^{\beta_e E t} - 1}.$$

Substituting $x = \beta_e E t$ we obtain

$$\frac{1}{x_1 - x_0} \int_{x_0}^{x_1} \frac{x^2 \, dx}{e^x - 1}.$$

The integral can be written as $-x^2 Li_1(e^{-x}) - 2x Li_2(e^{-x}) - 2Li_3(e^{-x})$ where Li_m are the polylogarithm functions. Figure 9(b) shows the result for a time interval $\beta_e t_0 = 0.375$, $\beta_e t_1 = 8.35$.

The similarity of figures 8 and 9(b) is striking and strongly supportive of the model for GRBs suggested in this paper.

More accurate pictures could be made by taking the exact formulae for $1 + z$ and ρ and performing numerical integration, but one should also bear in mind that the emitter temperature is in general a function of emitter time u and a short interval of receiver time near the beginning of the burst corresponds to a very long interval of emitter time.

There are many other features that observers identify for GRBs, e.g. spectral lag, variability, spectral peak photon energy, time of jetbreak and minimum rise time (e.g. [Scha]), which may also merit reinterpretation.

4.4 Directional track

One can also ask what is the track of the image of the emitter across the receiver's sky. It is an arc of a great circle, but the important question is how fast it moves in receiver time.

As in section 2.8, take the receiver geodesic to be $x_0 = \sinh t$, $x_1 = \cosh t$, $x_j = 0$ for $j = 2, 3, 4$, and the emitter geodesic to be $x_0 = a \sinh u$, $x_1 = c \cosh u$, $x_2 = s \cosh u$, $x_3 = b \sinh u$, $x_4 = 0$, where $a = \cosh \phi$, $b = \sinh \phi$, $c = \cos \theta$, $s = \sin \theta$.

Then the null geodesic comes to the receiver along the vector $(\sinh t - a \sinh u, \cosh t - c \cosh u, -s \cosh u, -b \sinh u, 0)$. The last three coordinates define the direction in the receiver's inertial frame from which it sees the emitter. Thus the receiver sees the emitter in the plane $x_4 = 0$, coming from an angle η relative to the x_2 direction with

$$\tan \eta = \frac{b}{s} \tanh u.$$

We can work out how fast the direction changes with receiver time:

$$\sec^2 \eta \frac{d\eta}{dt} = \frac{b}{s(1+z)} \operatorname{sech}^2 u,$$

using $dt/du = 1 + z$. Thus

$$\frac{d\eta}{dt} = \frac{sb}{1+z} \frac{\operatorname{sech}^2 u}{s^2 + b^2 \tanh^2 u}.$$

Expanding about the initial time $u = -\infty$, $t = t^*$, we already worked out that $1 + z \sim t - t^*$, thus $u \sim \log t - t^*$ and so sech $u \sim 2e^{-|u|} \sim 2(t - t^*)$. It follows that near the initial time,

$$\frac{d\eta}{dt} \sim \frac{4sb(t - t^*)}{s^2 + b^2}.$$

Thus the angular velocity of observation starts at zero and remains very small for $t - t^*$ short compared to the de Sitter time. The angular motion is particularly slow when θ is small or ϕ is large.

5 Beyond de Sitter space

5.1 Perturbing the metric on deS

We start by thinking about small perturbations of de Sitter space. Time-like geodesic flow in deS is Anosov [MR] so structurally stable; in particular, under small smooth perturbation to the metric every emitter-receiver pair is uniformly close to an unperturbed one.

As far as we can see, the structural stability result does not extend to the null geodesics, though it is possible that being limits of time-like ones, some uniform bounds might allow one to do so, in which case the whole picture of null geodesics between two time-like geodesics would be topologically stable to small perturbation. Still for the present we do not assume this.

The stability of our pair of time-like geodesics implies that if we construct a fibration of a neighbourhood of one of the geodesics by space-like leaves perpendicular to the geodesic then the perturbed geodesic intersects each leaf in a nearby point; the time-parametrisation is not necessarily preserved but the derivative of one with respect to the other is uniformly near 1.

We can suppose the receiver geodesic does not change because we can apply a diffeomorphism to move it and make a different perturbation of the metric to compensate. Suppose temporarily that the null geodesics to r do not change either (which of course is wrong in general).

The past light-cones of the receiver define a function τ on a big patch of space-time which is the receiver time t at which a null geodesic from a point can reach the receiver. In de Sitter space with the standard receiver geodesic this can be calculated to be $\tau = \log \frac{1+|y|}{x_1 - x_0}$ where y is short for (x_2, x_3, x_4). It is defined for $x_1 > x_0$. The sheets $\tau =$ constant form a fibration of a neighbourhood of the emitter geodesic by leaves transverse to any time-like curve and it follows from the Anosov property (as remarked above) that the perturbed emitter intersects each leaf in a single point and hence that the perturbed function $u \mapsto t$ is still monotonic.

However if we drop the assumption that null geodesics do not change then there is no reason to expect that $u \mapsto t$ should be monotonic. Under generic perturbation of the metric, past light cones develop Legendrian singularities (e.g. [FS, EN, EBD]). Indeed [EBD] estimate that a typical light-cone has 10^{22} caustics. The caustics for a single light-cone are of dimension 2, thus for a one-parameter family of light cones (say the backward light cones to the receiver geodesic), the caustic set is a subset of space-time of dimension 3. The emitter geodesic intersects this generically in isolated points.

We explain here the effect on the (t_e, t_r)-relation of passage of the emitter geodesic through a generic caustic of the receiver light-cone family (we use the original notation for emitter and receiver time to avoid confusion with u which will be used here as label for a light ray). From [FS], a normal form for a wavefront (a light-cone being an example) in a neighbourhood of a generic caustic is the image of the mapping

$$(u, v, \tau) \mapsto (-2u^3 + \tau, 3u^2 - \tau u, \tau\sqrt{1-u^2}, v)$$

into a local Minkowski space approximation $ds^2 = -dx_0^2 + \sum_{j=1}^3 dx_j^2$. Here the light rays are $(u, v) =$ constant and τ is an affine parameter along them. This has caustic given by $\tau = 6u(1 - u^2)$, hence it is the parametrised set of points of the form $(6u - 8u^3, 6u^4 - 3u^2, 6u(1-u^2)^{3/2}, v)$. As generic emitter intersecting the caustic we can take $x_j = 0$ for $j = 1, 2, 3$ with proper time x_0. It intersects the caustic at the origin. Rather than considering how the receiver light-cone displaces as the receiver time changes, we consider the effect to be equivalent to a displacement of the emitter geodesic. As typical displacement we take $x_1 = x = kt_r$, $x_j = 0$ for $j = 2, 3$, with proper time $t_e = x_0 + ct_r$ for some constants $c, k > 0$. This intersects the wavefront in points given by $\tau = 0$, $x_0 = -2u^3$, $x_1 = 3u^2$. Thus the intersections correspond to $u = \pm\sqrt{x/3}$ for $x \geq 0$ and they have $x_0 = \mp 2(x/3)^{3/2}$. Thus $t_e = ct_r \mp 2(kt_r/3)^{3/2}$, giving a cusp in the (t_e, t_r)-diagram with slope c. To determine the received flux, we need the corrected luminosity distance ρ. This passes through 0 for the case that the emitter intersects the caustic. From [Per] one can deduce that ρ is proportional to the square root of affine distance along the light ray from the emitter to the caustic. We take x_0 as affine parameter. For the light ray labelled by $u = -\sqrt{x/3}$ and $v = 0$, x_0 goes from $2(x/3)^{3/2}$ at the emitter to $(8x/3 - 6)\sqrt{x/3}$ at the caustic. Thus the change in affine parameter is $6(x/3 - 1)\sqrt{x/3}$, which is to leading order $-6\sqrt{x/3}$. Thus ρ is proportional to $t_r^{1/4}$. The redshift is roughly constant at $z = 1/c - 1$. This gives received flux $\Phi = P/((1+z)\rho)^2$ proportional to $t_r^{-1/2}$.

Thus the received intensity is infinite at the time of the cusp, but it is an integrable singularity, so the received energy in a time interval around 0 is finite. The redshift does nothing exceptional. The effect is to make an additional spike on the light-curve, which could explain some more of the structure observed. It is a relativistic version of the twinkling of stars. The spikes can be forward or backward facing, according to the orientation of the cusp. The generic first singularity of a wavefront is a cusp (as in Fig.1(a) of [EBD]), generating a pair of folds. The pair of folds produce a (t_e, t_r) diagram like Figure 10, and thus a forward and backward pair of inverse square root spikes.

One can think of the additional spikes as analogous to 'echoes' of the main GRB.

Note that by the relativistic version of Fermat's principle formulated by Kovner (e.g. [Per]), light rays from a time-like curve (call it the emitter) to a point of space-time are critical points of the emission time in the space of light-like paths from the emitter to the receiver point. It follows that (u, t) curves can not have vertical tangents (nor by time-reversal, horizontal ones). Also note

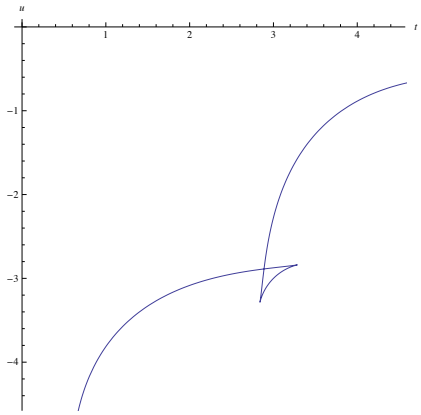

Figure 10. (t_e, t_r)-relation for an emitter intersecting the backward light-cone beyond a cusp.

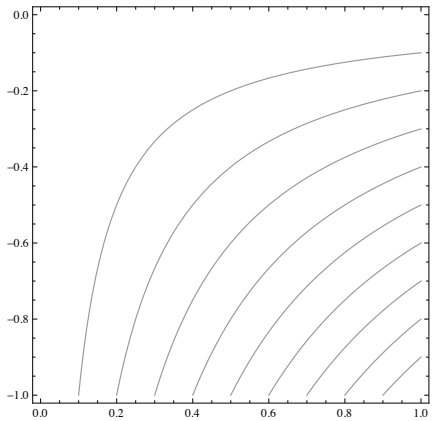

Figure 11. (t_e, t_r)-relation in the presence of a black hole.

that the ray that leaves the emitter latest has no conjugate points and is called the “primary ray". Other images of the emitter visible at the same receiver time have come from earlier emitter time and followed paths with at least one conjugate point.

If large perturbation from de Sitter space is considered then larger effects can be expected. For example, on introducing a Schwarzschild black hole, as in Kottler space, the (u, t)-diagram gains an infinite series of curves, as sketched in Figure 11, corresponding to light rays making different numbers of turns around the black hole. The added travel time per turn in the black hole's frame is asymptotically $6\pi\sqrt{3}M$ for black hole mass M (the Schwarzschild time to perform one revolution around the light-sphere $r = 3M$). Successive curves are presumably fainter.

Images of the emitter via different rays in general come in to the receiver from different directions. For example, the two images from passage through a generic caustic separate in direction like square root of receiver time, but the coefficient depends on the situation.

5.2 Comparison with Friedmann universes

Finally we make a comparison with what one would see in a Friedmann universe.

The standard cosmological model is a Friedmann universe, with metric

$$ds^2 = -dt^2 + S(t)^2|dx|^2$$

for some scale factor $S(t)$ going to 0 as time t goes to the big bang, which we place at $t = 0$, and Euclidean metric on the space-slices. If we are on a Hubble flow line $\mathbf{x} =$ constant then we see every time-like geodesic redshifted. There is a first time $t^* > 0$ we begin to see it, corresponding to emitter time 0. We see it infinitely redshifted, unless its velocity is directed exactly towards us

in which case it is finitely redshifted. We include the calculations, because the only case we have seen treated is that of textbooks, where both receiver and emitter are along lines of the Hubble flow.

Firstly, each null geodesic lives in a plane spanned by the time direction and a spatial vector, without loss of generality the (t,x)-plane. It satisfies $dx/dt = 1/S(t)$, thus it travels a displacement $\Delta x = \int dt/S(t)$.

Similarly, each time-like geodesic lives in a 2-plane, without loss of generality again the (t,x)-plane, though in general it is different from those for the null geodesics we will want to consider. Working out the Christoffel symbols for the Friedmann metric, the equations for geodesics boil down to

$$x_{\tau\tau} + 2\frac{\dot S}{S}x_\tau t_\tau = 0 \tag{5.1}$$

$$t_{\tau\tau} - S\dot S x_\tau^2 = 0, \tag{5.2}$$

where τ is proper time, i.e. such that the conserved quantity $-t_\tau^2 + S^2x_\tau^2 = -1$ and future-pointing. Write $v = x_\tau$ so $t_\tau = \sqrt{1+S^2v^2}$ and

$$v_\tau = -2\frac{\dot S}{S}v\sqrt{1+S^2v^2}.$$

From these we deduce that $v_t = -2v\dot S/S$, so $v = k/S^2$ for some constant k, $t_\tau = \sqrt{1+k^2/S^2}$ and $x_t = k/\sqrt{S^4+k^2S^2}$. The constant k is zero iff the geodesic is on the Hubble flow.

Introduce new time coordinate $T = \int_0^t dt/S(t)$, so the null geodesics are $x = x_0 + T$ (in their own planes, of course), and for the time-like geodesics $x_T = k/\sqrt{S^2+k^2}$ and $\tau_T = S^2/\sqrt{S^2+k^2}$.

Taking receiver geodesic along $\mathbf{x} = \mathbf{A}$, the light from the emitter at (x,T) arrives at

$$T_r = T + |x - A|. \tag{5.3}$$

Proper time on the receiver is just t. Thus the first time t^* of receipt of light from the emitter is defined by $\int_0^{t^*} dt/S(t) = |x(0) - A|$. Weyl should have objected to Friedmann universes!

Differentiating (5.3) with respect to emitter proper time τ we obtain

$$\frac{dT_r}{d\tau} = \frac{dT}{d\tau} + \langle\frac{dx}{d\tau}, x - A\rangle/|x - A|.$$

Writing θ for the angle between the velocity of the emitter and the vector $x - A$ and using the above equations, we end up with

$$1 + z = \frac{dt_r}{d\tau} = \frac{S(t_r)}{S(t_e)^2}\left(\sqrt{S(t_e)^2 + k^2} + k\cos\theta\right).$$

Since $S(t_e)$ starts at 0, we obtain infinite redshift initially, except if $\cos\theta = -1$ when we obtain $1 + z = 1/(2k)$ initially.

One could work out the (u,t)-relation for general receiver geodesic as well as emitter geodesic, in essentially the same way.

If one allows the case $S(t) = e^t$, however, for which $S(t) \to 0$ as $t \to -\infty$ instead of in finite time, then the Friedmann universe is one half of de Sitter space and our scenario for gamma-ray bursts applies.

We believe there is room between de Sitter space and big bang universes for our mechanism for gamma-ray bursts to apply.

References

[BAT] Barthelmy, SD et al, The Burst Alert Telescope (BAT) on the SWIFT Midex Mission, Space Science Reviews 120:3–4 (2005) 143–164, arXiv:astro-ph/0507410

[EBD] Ellis GFR, Bassett BACC, Dunsby PKS, Lensing and caustic effects on cosmological distances, Class. Quantum Grav. 15 (1998) 2345–2361
doi:10.1088/0264-9381/15/8/015

[EN] Ehlers J, Newman ET, The theory of caustics and wavefront singularities with physical applications, J. Math. Phys. 41 (2000) 3344–3378
doi:10.1063/1.533316

[FS] Friedrich H, Stewart JM, Characteristic initial data and wave front singularities in general relativity, Proc. Roy. Soc. Lond. A 385 (1983) 345–371
doi:10.1098/rspa.1983.0018

[Gol] Goldstein A et al, The Fermi GBM Gamma-ray burst spectral catalog; the first two years, Astrop J Suppl 199 (2012) 19–45
doi:10.1088/0067-0049/199/1/19

[Gr] Greiner JC, Localized GRBs, www.mpe.mpg.de/∼jcg/grbgen.html

[GP] Guetta D and Piran T, The BATSE-Swift luminosity and redshift distributions of short-duration GRBs, Astronomy and Astrophysics 453 (2006) 823–828
arXiv:astro-ph/0511239

[K+] Kouveliotou C, Meegan CA, Fishman GJ, Bhat NP, Briggs MS, Koshut TM, Paciesas WS, Pendleton GN, Identification of two classes of gamma-ray bursts, Astrop J 413 (1993) L101–4

[LC] Levi-Civita T, Realtà fisica di alcuni spazî normali del Bianchi, Rendiconti Reale Accademia Dei Lincei 26 (1917) 519–531

[MR] MacKay RS, Rourke C, Natural observer fields and redshift, J. Cosmology 15 (2011) 6079–6099
http://msp.warwick.ac.uk/~cpr/paradigm/redshift-nat-final.pdf

[MR2] MacKay RS, Rourke C, Natural flat observer fields in spherically symmetric spacetimes, J Phys A 48 (2015) 225204 (11pp)

[Mes] Meszaros P, Gamma-ray bursts, Rep Prog Phys 69 (2006) 2259–2321

[MTW] C W Misner CW, Thorne KS, Wheeler JA, Gravitation, Freeman (1973)

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

[NP] Newman ET, Penrose R, An Approach to Gravitational Radiation by a Method of Spin Coefficients, J. Math. Phys. 3 (1962) 566–578
doi:10.1063/1.1724257

[Per] Perlick V, Gravitational lensing from a spacetime perspective, Living Reviews Relativity 7 (2004) 9

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

[S] Schrödinger E, Expanding universes, CUP (1956)

[Scha] Schaeffer BE, The hubble diagram to redshift $>$ 6 from 60 gamma-ray bursts, Astrop J 660 (2007) 16–46

[deS17] Sitter W de, On the relativity of inertia, Kon. Ned. Acad. Wet. Proc. 19 II (1917) 1217–1225

[deS18] Sitter W de, On the curvature of space, Kon. Ned. Acad. Wet. Proc. 20 I (1918) 229–243

[We] Weyl H, Zur allgemeinen Relativitätstheorie, Phys Zeits 24 (1923) 230–232

[We30] Weyl H, Redshift and relativistic cosmology, Phil Mag [7] 9 (1930) 936–943

Author information

Robert S MacKay, Colin Rourke, Mathematics Institute University of Warwick, Coventry CV4 7AL, United Kingdom.
E-mail: R.S.MacKay@warwick.ac.uk, cpr@msp.warwick.ac.uk

Received: February 14, 2015.

Accepted: May 17, 2015