Flares in gamma-ray burst X-ray afterglows as prompt emission from slightly misaligned structured jets

Raphaël Duque, Paz Beniamini, Frédéric Daigne, Robert Mochkovitch

Flares in gamma-ray burst X-ray afterglows as prompt emission from slightly misaligned structured jets

Raphaël Duque [1,2]★ Paz Beniamini,[3,4,5] Frédéric Daigne[1] and Robert Mochkovitch[1]

[1]*Institut d'Astrophysique de Paris, UMR 7095, CNRS, Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France*
[2]*Institut für Theoretische Physik, Goethe Universität Frankfurt am Main, D-60323 Frankfurt am Main, Germany*
[3]*Department of Natural Sciences, Open University of Israel, 1 University Road, 43107 Ra'anana, Israel*
[4]*Astrophysics Research Center of the Open university (ARCO), The Open University of Israel, P.O. Box 808, Ra'anana 43537, Israel*
[5]*Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*

ABSTRACT

We develop a model to explain the flaring activity in gamma-ray burst X-ray afterglows within the framework of slightly misaligned observers to structured jets. We suggest that flares could be the manifestation of prompt dissipation within the core of the jet, appearing to a misaligned observer in the X-ray band because of less favourable Doppler boosting. These flares appear during the afterglow phase because of core–observer light travel delays. In this picture, the prompt emission recorded by this observer comes from material along their line of sight, in the lateral structure of the jet, outside the jet's core. We start by laying down the basic analytical framework to determine the flares characteristics as a function of those of the gamma-ray pulse an aligned observer would see. We show that there is viable parameter space to explain flares with typical observing times and luminosities. We then analytically explore this model, showing that it naturally produces flares with small aspect ratios, as observed. We perform fits of our model to two *Swift*/XRT flares representing two different types of morphology, to show that our model can capture both. The ejection time of the core jet material responsible of the flare is a critical parameter. While it always remains small compared to the observed time of the flare, confirming that our model does not require very late central engine activity, late ejection times are strongly favoured, sometimes larger than the observed duration of the parent gamma-ray burst's prompt emission as measured by T_{90}.

Key words: galaxies: jets – radiation mechanisms: general – *(stars:)* gamma-ray burst: general – *(stars:)* gamma-ray burst: individual: GRB060719 – *(stars:)* gamma-ray burst: individual: GRB100816A – X-rays: bursts.

1 INTRODUCTION

Flares are sudden rebrightenings in the afterglow phases of gamma-ray bursts (GRBs), observed primarily in the X-ray band (Burrows et al. 2005b; Nousek et al. 2006), though also in the optical (Li et al. 2012; Swenson et al. 2013; Yi et al. 2017) and radio bands (e.g. Malesani et al. 2007). Flares occur in around one third of observed GRB afterglows, and most observed flares occur less than 1000 s after the prompt trigger (e.g. Chincarini et al. 2010; Yi et al. 2016). From the first catalogues of flares in *Swift*/XRT light curves, flares showed the salient feature of having small and tightly distributed *aspect ratios* (i.e. the ratio of their width Δt to their arrival time t; Chincarini et al. 2007). It was also noted that flare morphologies were quite diverse – with both fast and slow rising and decay phases – and that they mimicked GRB prompt pulses, with a very similar distribution of rise-to-decay-time ratios (Chincarini et al. 2010). Flares exhibit other features remarkably analogous to prompt pulses such as spectral lags, lag–luminosity correlations, and width–energy band relations (Margutti et al. 2010). Pulse temporal profile and variability time-scale analyses of prompt emission and X-ray flares also find common traits for both phenomena (Sonbas et al. 2013; Guidorzi et al. 2015).

These temporal and spectral similarities with prompt pulses suggest a common origin for flares and prompt emission. For example, the central engine of GRBs could have a second episode of activity, explaining the delay between prompt emission and X-ray flares (e.g. Burrows et al. 2005b; Fan & Wei 2005; Liang et al. 2006; Lazzati & Perna 2007). Causes for the engine's delayed restarting can, for example, be fragmentation and accretion of a collapsing star (King et al. 2005), instability-induced variability in accretion around the central object (Perna, Armitage & Zhang 2006), or magnetic activity of the young pulsar produced by the merger in non-collapsar GRBs (Dai et al. 2006). However, in these late-engine-activity models, the emission in the X-ray rather than the gamma-ray bands requires an explanation. Furthermore, producing small aspect ratios often requires tuning the second activity's duration to the time of quiescence between the two episodes in an unnatural way. Finally, in some cases, the energy requirements for powering the X-ray flares are often incompatible with the energy available *at that time* (Beniamini & Kumar 2016), from, e.g. a rapidly spinning magnetar – a widely invoked model for GRB central engines. Alternatively, the central engine can have a single episode of ejection, in which case the dissipation of energy in the jet can be delayed, as in delayed magnetic dissipation or internal shocks (Giannios 2006; Yu & Dai 2009; Troja et al. 2015; Pescalli et al. 2018). Other models for flares suggest

★ E-mail: duque@physik.uni-frankfurt.de

an origin distinct from that of prompt dissipation, such as emission from the reverse shock propagating in a stratified ejecta (Hascoët et al. 2017; Lamberts & Daigne 2018; Ayache, van Eerten & Daigne 2020), Compton up-scattering of photons from the reverse shock when crossing the forward shock (Kobayashi et al. 2007) or from the forward shock on a preceding relativistic outflow (Panaitescu 2008), or photospheric emission from material moving with modest Lorentz factors compared to that producing the prompt, but ejected at roughly the same time (Beniamini & Kumar 2016).

In a recent publication (Beniamini et al. 2020a), we developed a model to explain the emergence of plateaus in some GRB afterglows, another remarkable feature of GRB afterglows consisting of extended phases of near-flat flux evolution occurring before the regular decay of afterglows (Nousek et al. 2006). This model, initially suggested by Eichler & Granot (2006), is based on a set-up where the GRB jet possesses angular structure and the observer is slightly misaligned with the GRB jet, to within a few times the jet's core half-opening angle, i.e. $\theta_v \lesssim 2\theta_j$. It is indeed likely that this regime of viewing angles is that assumed when detecting GRBs observed at cosmological distances (Beniamini & Nakar 2019). In this model, the shallow flux evolution of the plateau is produced by the fact that any region in the decelerating jet is only revealed to the observer when this region has slowed down enough for the line of sight to be included in the material's beaming cone. Thus, as the jet decelerates, the observer progressively discovers material closer to the jet's core. Because of the jet structure, this material is intrinsically brighter but less boosted because it is pointed further away from the observer. This can lead to a shallow evolution of the total afterglow flux. In Beniamini et al. (2020a), we derived analytically the duration and flux of plateaus expected in this picture as a function of the jet structure and the observer's viewing angle. We showed how established correlations between plateau duration and flux level and the parent GRB's prompt properties naturally arise in this model.

Recently, GRB jet structure has become an increasingly important ingredient in GRB afterglow modelling (e.g. Ascenzi et al. 2020; Beniamini, Granot & Gill 2020b; Oganesyan et al. 2020; Lamb et al. 2021; Takahashi & Ioka 2021), especially since the historical insight provided by the multimessenger analysis of the outflow from GW170817 (Mooley et al. 2018; Ghirlanda et al. 2019). In this paper, we set out to interpret flares in GRB afterglows within the same physical set-up as our above-mentioned plateau model: slightly misaligned lines of sight to a structured jet. Acknowledging the aforementioned similarities between X-ray flares and GRB prompt pulses, we also posit a common origin for the two. However, we will explain the delayed occurrence of the flares not by their delayed emission, but rather by the light traveltime between the flare production site within the core and the misaligned observer: We suggest that flares in GRB X-ray afterglows are the manifestation of prompt dissipation in the core of the jet, as seen from slightly off-axis lines of sight. Because of relativistic effects, this radiation appears delayed, dimmer, and downshifted in energy. In other words, X-ray flares are *deboosted* versions of gamma-ray pulses from prompt energy dissipation in the core.

We present our unified picture for plateaus and flares in Fig. 1: For an aligned viewer (A), the prompt emission comes from the core jet shining in gamma-rays (green), and the afterglow phase contains the early steep decay (ESD) and radiation from the decelerating forward shock (grey); all other jet regions are too weak and not boosted enough to contribute to the aligned observer's signal. For a misaligned observer (B), the prompt emission and ESD come from the material down their line of sight (red). Progressively the structured jet decelerates, giving rise to the plateau phase (grey). In
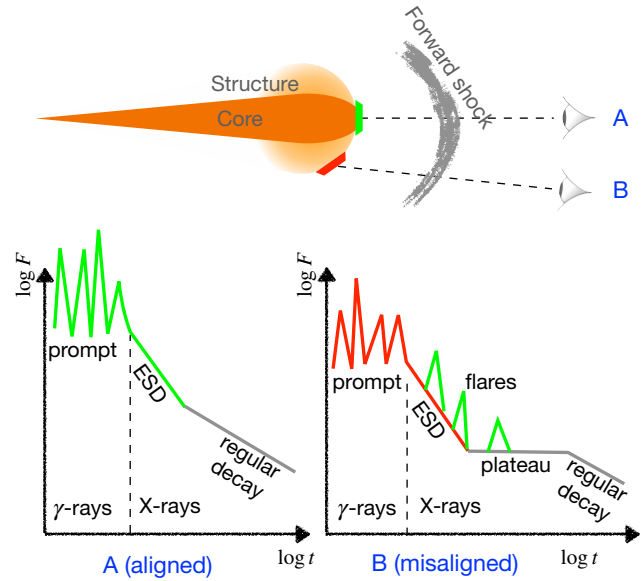

Figure 1. Schematic description of our unified picture for flares and plateaus. The colours of the elements of the light curves correspond with the colours of the emitting regions in the jet: core (green) or lateral structure (red). Each component of the prompt and afterglow phases of aligned (A) and misaligned (B) viewers comes from a different region, according to our picture. For the misaligned observer, the emission from the core matter appears as flares in the X-ray band, atop the ESD and the plateau phase.

the mean time, prompt photons from the core (green) travel to the observer, and reach them as X-ray flares, i.e. deboosted and dimmer than photons that would have reached the aligned observer.

Note that we consider a single central engine activity episode, and the delay in flare occurrence is a geometrical effect. We anticipate Section 2 in mentioning that the arrival time for flares in such a picture is bounded to $t_{\rm flares} \lesssim 1000$ s after the prompt trigger. We are therefore dealing with *early flares*. Remarkably, these early flares seem to constitute a distinct (and largely statistically dominant; Swenson & Roming 2014) class of flares, as shown by their distinctive post-peak decay slopes and distribution of flare-to-continuum contrasts ($\Delta F/F$) with respect to late flares ($t_{\rm flare} \gtrsim 1000$ s, Bernardini et al. 2011; Margutti et al. 2011). While the dichotomy in $\Delta F/F$ could result from the late flares most often occurring during the plateau phase of the afterglow and the early flares during the ESD, the dichotomy in decay slopes suggests a different origin for these two classes, and our model's natural restriction to the early class is a further motivation to explore this picture.

We outline this new framework for flares and make a first exploration of parameter space in Section 2. We study the conditions for flares to appear above the underlying afterglow continuum in Section 3, where we also expose a feature of our model by which flares that rise above the afterglow continuum tend to be thinner. In Section 4, we confront our model to *Swift* X-ray data by making fits to light curves of actual GRB afterglows with flares. We summarize our results and discuss the underlying prompt dissipation mechanism and consequences of our model in Section 5.

2 MODEL OUTLINE AND BASIC PROPERTIES

We consider a shell of ultra-relativistic matter with Lorentz factor Γ ejected at a time $t_{\rm ej}$ from the central engine, within the core of the jet. At an emission time $t_{\rm e}$, this shell reaches a dissipation radius $R_{\rm e} =$

$\beta(t_e - t_{ej})$ and radiates energy which, for a core-aligned observer appears as gamma-rays. For this aligned observer, this radiation is observed at time

$$t_{on} = t_{ej} + \frac{1-\beta}{\beta} \frac{R_e}{c} \tag{1}$$

where $\beta \sim 1$ is the shell's velocity.

We now consider a misaligned observer, lying at a viewing angle θ_v from the jet's axis, with $\theta_v > \theta_j$, where θ_j is the core's half-opening angle. For this observer, the first photons from this shell's radiation arrive at a time $t_{off} = t_{ej} + \frac{1-\beta\cos(\theta_v - \theta_j)}{\beta} \frac{R_e}{c}$. Stated differently

$$t_{off} = \mathcal{S}t_{on} - (\mathcal{S} - 1)t_{ej}, \tag{2}$$

where we have denoted

$$\mathcal{S} \equiv \frac{1 - \beta\cos(\theta_v - \theta_j)}{1 - \beta}, \tag{3}$$

which we will refer to as the *stretch factor*. This factor is the ratio of the Doppler boosts between the aligned and misaligned observers. In the ultra-relativistic limit ($\Gamma \gg 1$) and for slightly off-axis observers ($\theta_v - \theta_j \ll 1$), the stretch factor reduces to $\mathcal{S} \simeq 1 + \Gamma^2 (\theta_v - \theta_j)^2$.

Furthermore, the peak energy of the detected spectrum transforms as $\mathcal{S}^{-1}$ between the two observers' rest frames:

$$e_{off} = \mathcal{S}^{-1} e_{on} \tag{4}$$

Finally, considering a slightly misaligned line of sight ($\theta_j < \theta_v \lesssim 2\theta_j$) we show in Appendix A that the peak bolometric luminosity of the detected radiation transforms according to (equation A6):

$$L_{off} = f_{geo}\mathcal{S}^{-3}L_{on}, \tag{5}$$

where $f_{geo} \leq 1/2$ is a numerical factor accounting for the transverse angular size of the core jet as seen from the slightly off-axis line of sight. For an observer on the edge of the core, $f_{geo} = 1/2$, and f_{geo} decreases with larger viewing angle. In Beniamini et al. (2020a), we considered the observer to be placed at $\theta_v/\theta_j \sim 1.3$ to illustrate the results of our plateau model. In our numerical exploration of the present model (Section 3.3), we shall adopt this same geometric configuration. In this case, we find that equation (5) is best reproduced with $f_{geo} \sim 0.14$.

These relations show that what is seen as a prompt pulse by an observer aligned with the jet core transforms, for a slightly misaligned observer, into a signal that is delayed, softer, and less luminous. In this picture, we will use the term *pulse* for the signal detected by a core-aligned observer, and *flare* for the deboosted emission a misaligned observer detects. A gamma-ray pulse turns into an X-ray flare when seen by an off-axis observer.

For example, let us consider $\Gamma = 150$ and $\theta_v - \theta_j = 0.03\,\mathrm{rad} \sim 5/\Gamma$, a viewing angle that could typically lead to plateau behaviour in the afterglow once the structure decelerates, according to the misaligned-observer interpretation developed in Beniamini et al. (2020a; see their fig. 1). This configuration leads to $\mathcal{S} = 21$. We take the following typical GRB pulse characteristics: a pulse observed at $t_{on} = 8\,\mathrm{s}$, coming from a shell ejected at $t_{ej} = 3\,\mathrm{s}$ with peak bolometric luminosity $L_{on} = 5 \times 10^{52}\,\mathrm{erg\,s^{-1}}$ and peak energy $e_{on} = 300\,\mathrm{keV}$. On the slightly misaligned line of sight, such a pulse would appear as an afterglow rebrightening occurring at $t = 110\,\mathrm{s}$ with a peak energy of $e_{off} = 14\,\mathrm{keV}$, in the X-ray band. The X-ray luminosity – or bolometric, as the signal peaks in the X-rays – of this flare is $L_{off} = 3 \times 10^{48}\,\mathrm{erg\,s^{-1}}$. It seems therefore that this mechanism can explain the flaring activity in GRB afterglows.

Let us generalize this order-of-magnitude estimate. We denote by L_{flare} the peak bolometric luminosity of an observed flare and t_{flare}

its peak time. We seek conditions under which this flare can be interpreted as deboosted core prompt emission. We denote by $\tau = R_e/2\Gamma^2 c$ the shell's angular time-scale. Assuming the duration of the emission is negligible – i.e. instantaneous dissipation of energy in the shell, the duration of the pulse as seen by an aligned observer is well approximated by the delay between the arrivals of photons from their beaming cone's axis and edge:

$$\Delta t_{on} = \frac{R_e}{c} (1 - \cos 1/\Gamma) \tag{6}$$

$$\sim \frac{R_e}{2c\Gamma^2} \tag{7}$$

$$= \tau. \tag{8}$$

Assuming the dissipation is instantaneous simplifies much of the derivation below. We will adopt this hypothesis as this is a first exploration of this model and it allows carrying out the analytical development further. In practice, the relationship between τ and Δt_{on} depends on the prompt dissipation mechanism. Different mechanisms would affect the light-curve profiles of the flares we will obtain below, but not the general features of the model. In Section 5.2, we will discuss this point related to GRB emission mechanisms in more detail.

Therefore, writing $\Delta t_{on} = \tau \sim (1 - \beta)R_e/c$ and using equations (1) and (2), we arrive at this new form for the flare peak time:

$$t_{flare} = \tau S + t_{ej}, \tag{9}$$

where we used $\Delta t_{on} = t_{on} - t_{ej}$.

Furthermore, the bolometric luminosity observed by the aligned observer is approximately $L_{on} \sim E_{iso}/\Delta t_{on}$, where E_{iso} is the isotropic-equivalent source-frame dissipated energy in the shell responsible for the emission of the pulse. Using equation (5), we finally obtain the luminosity of the corresponding flare as seen by a slightly off-axis observer:

$$L_{flare} = f_{geo}S^{-3}\frac{E_{iso}}{\tau}. \tag{10}$$

In our picture, the shell's ejection occurs during the central engine activity, which lasts for a duration T_{CE} depending on the smaller scale physics around the central engine and in the accretion disc. As we do not have a firm constraint on this duration T_{CE}, we shall consider it as a control parameter in the model fits we carry out in Section 4. Indeed, T_{CE} should not be much larger than the duration of GRB prompt phases, often estimated by T_{90}. Similarly, the durations of pulses in GRBs are generally less than a few seconds (e.g. Hakkila et al. 2018). Therefore, the Δt_{on} (or τ) we can consider are also constrained, to being less than a few seconds. Finally, for the prompt gamma-ray pulses to transform to X-ray flares, the stretch factor must typically be $\mathcal{S} \lesssim 100$, as mentioned in the numerical example above. As we shall soon see, keeping reasonable values for E_{iso} also implies that $\mathcal{S} \lesssim 100$, due to the flare flux suppression by $\mathcal{S}^{-3}$ (equation 5). Therefore, we conclude from equation (9) that the flares explained in our off-axis mechanism cannot appear much later than $t_{flare} \sim 1000\,\mathrm{s}$. As mentioned in Section 1, there seems to be a dichotomy both in post-peak decay slopes and in the $\Delta F/F$ distribution between early ($t_{flare} < 1000\,\mathrm{s}$) and late ($t_{flare} > 1000\,\mathrm{s}$) flares (Bernardini et al. 2011). Early flares may thus be a distinct subclass of flares produced by a specific mechanism and the natural production of such arrival times by our model further motivates to pursue this picture. Moreover, these early flares are by far the most numerous in the observed population: In the source rest frame (i.e. redshift-corrected), 80 per cent of X-ray flares occur less than 260 s after the prompt trigger (Yi et al. 2016);
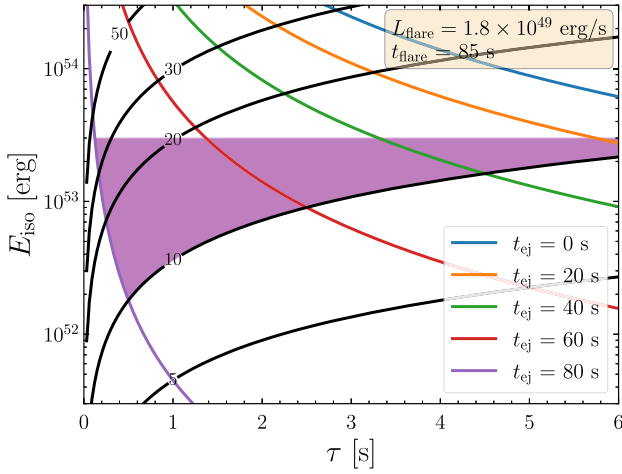
Figure 2. Misaligned structured-jet flare model parameter space exploration to reproduce the median XRT flare. The coloured lines represent the $E_{\rm iso}$ energy required to produce a typical flare peaking at $t_{\rm flare} = 85$ s and with peak luminosity $L_{\rm flare} = 1.8 \times 10^{49}$ erg s^{-1} as a function of the shell's geometrical time-scale τ and for different ejections times from 0 to 80 s, an estimate of GRB central engine activity duration. The black lines are contours of the corresponding required $\mathcal{S}$-factor, values are indicated on the lines. The region in this parameter space consistent with the off-axis geometrical set-up of our model and with expected ranges for shell dissipated energies and GRB pulse durations is marked in purple. We find that there is available parameter space for our model to explain the typical flare.

see however in Section 5.4 a discussion on the nature of these early flares and possible pollution by prompt emission.

The typical values required for $\mathcal{S} \lesssim 100$ in our picture confirm that our flare model applies only to slightly misaligned observers. Indeed, for large Lorentz factors this condition translates to $\theta_j < \theta_v \lesssim \theta_j + \frac{10}{\Gamma}$. For $\Gamma \geq 100$, this is $\theta_v - \theta_j \leq 10/\Gamma \leq 0.1 \lesssim \theta_j$.

Inverting equations (9) and equations (10), we conclude that, for observed flare properties $L_{\rm flare}$ and $t_{\rm flare}$ and assuming the shell's ejection time $t_{\rm ej}$, the isotropic-equivalent dissipated energy required to produce the flare is given as a function of τ by

$$E_{\rm iso}(\tau) = \tau^{-2} \frac{L_{\rm flare}}{f_{\rm geo}} \left(t_{\rm flare} - t_{\rm ej} \right)^3 . \tag{11}$$

Assuming these $E_{\rm iso}$ and τ indeed reproduce the flare, the corresponding stretch factor is

$$\mathcal{S} = \left(\frac{\tau L_{\rm flare}}{f_{\rm geo} E_{\rm iso}} \right)^{-1/3} . \tag{12}$$

In Fig. 2 (coloured lines), we plot $E_{\rm iso}(\tau)$ assuming ejections times from 0 to $T_{\rm CE} = 80$ s, an estimate of GRB central engine activity duration we deduce from the average duration of *Fermi*/GBM long bursts ($\overline{T_{90}} = 40$ s, Von Kienlin et al. 2020). Also, we assumed $f_{\rm geo} = 1/2$. In Fig. 2, we chose the median source-frame (i.e. redshift-corrected) $t_{\rm flare}$ and $L_{\rm flare}$ in the *Swift*/XRT afterglow flare sample: $\overline{t_{\rm flare}} = 85$ s and $\overline{L_{\rm flare}} = 1.8 \times 10^{49}$ erg s^{-1} (Chincarini et al. 2010; Yi et al. 2016). We also plotted isocontours of the corresponding $\mathcal{S}$-factor. This figure allows us to outline which model parameter values are required to reproduce the typical X-ray flare and if these values are consistent with our physical set-up and typical GRB quantities. According to the above discussion, the parameter space for our model is constrained by the following conditions:

(i) Shell ejection must occur during the primary central engine activity, thus $t_{\rm ej} \lesssim T_{\rm CE}$;

(ii) We must maintain a slightly misaligned line of sight, thus $\mathcal{S} \gtrsim 10$;

(iii) Prompt gamma-rays must be deboosted to the X-ray band, thus $e_{\rm on}/e_{\rm off} = \mathcal{S} \lesssim 100$;

(iv) An aligned observer should detect a classical prompt pulse with duration typically between 0.1 and 5 s: $0.1\,{\rm s} \lesssim \Delta t_{\rm on} \sim \tau \lesssim 10\,{\rm s}$;

(v) The shell must dissipate an energy in line with what is observed for single pulses in typical GRBs. For entire GRBs (as opposed to single pulses within a GRB light curve), the dissipated energies are in the range $E_{\rm iso}^{\rm GRB} = 10^{53 \pm 1}$ erg (e.g. Amati 2006). Assuming a few pulses per GRB, we can estimate that each pulse dissipates an energy on the order of $E_{\rm iso} = 10^{52.3 \pm 1}$ erg. We can thus consider this range for the shell's single dissipation episode.

The region in parameter space respecting all of these conditions is coloured in purple in Fig. 2: for any $(\tau, E_{\rm iso})$ pair in this zone, there corresponds an ejection time $t_{\rm ej}$ during the central engine activity and an $\mathcal{S}$-factor of a misaligned line of sight such that the flare appears at $\overline{t_{\rm flare}}$ with peak luminosity $\overline{L_{\rm flare}}$. This figure shows that there is open parameter space available to our model to explain typical flares as deboosted core prompt dissipation. Of course, there is much parameter space available for late ejection times ($t_{\rm ej} \sim T_{\rm CE}$), such that the arrival time of the flare is mainly due to delayed activity. More interestingly, there is also space with larger τ and smaller ejection times ($t_{\rm ej} \leq 60$ s), where the light traveltime from the core to the misaligned observer plays a role in the flare arrival time.

One may use plots like Fig. 2 to find solutions to flares in actual GRB afterglows, adapting $t_{\rm flare}$ and $L_{\rm flare}$. For brighter or later flares, equation (11) shows that larger $E_{\rm iso}$ are required, and the available parameter space will shrink. This equation also shows that allowing for larger ejection times eases the constraint on $E_{\rm iso}$. In Fig. 2 we bounded $t_{\rm ej}$ with a generic estimate for central engine activity duration, $t{\rm ej} \leq T_{\rm CE} \lesssim 80$ s. In fact, this bound should be on the duration of central engine activity *in directions within the core*, $T_{\rm CE}^{\rm c}$. This duration may be different than the duration of the central engine activity on the misaligned observer's line of sight $T_{\rm CE}^{\rm LOS}$, measured by the actual GRB in which the flare we seek to explain appeared. Here, we will suppose that the central engine activity has the same duration in all directions and discuss this hypothesis in Section 5.2.

For a given flare, one must also consider the actual temporal profile of the flare and seek solutions within this available parameter space that correctly fit the light curve. We will do so for two typical flare shapes in Section 4.

3 EARLY FLARE VISIBILITY

3.1 Conditions for flare visibility

The majority of X-ray flares occur early in the X-ray afterglow, during the ESD phase. Within the context of a structured jet, we will now outline the conditions for the flares produced by our mechanism to appear above the ESD.

We suppose the observer lies at an angle θ_v from the jet axis. For this observer, the ESD will be produced by high-latitude emission from the last shell that flashed on their line of sight. Introducing the dissipated energy $E_{\rm iso}^{\rm LOS}$ of this last shell and its angular time-scale $\tau^{\rm LOS}$, this ESD phase will have an approximate isotropic-equivalent bolometric luminosity of

$$L_{\rm ESD}^{\rm LOS}(t) = \frac{E_{\rm iso}^{\rm LOS}}{\tau^{\rm LOS}} \left(\frac{t}{\tau^{\rm LOS}} \right)^{-3} \tag{13}$$

for times larger than the end of the prompt emission.

Using the same notations as in Section 2 adding the superscript ^{c} for the shell in the core responsible for the flare, the flare's peak luminosity is still $L_{flare} = f_{geo}\mathcal{S}^{-3}E_{iso}^c/\tau^c$ and the arrival time $t_{flare} = \mathcal{S}\tau^c + t_{ej}$. We now introduce the jet structure, prescribing that, on average, the dissipated energies of shells in the core and on the θ_v line of sight are linked by[1]

$$\frac{E_{iso}^{LOS}}{E_{iso}^c} = \left(\frac{\theta_v}{\theta_j}\right)^{-a}, \tag{14}$$

where we chose a power-law for the energy structure for ease of the analytic development that will follow.

The Lorentz factor of the material in the jet also possesses structure. We adopt a similar power-law dependence for the material's average Lorentz factor:

$$\frac{\Gamma^{LOS} - 1}{\Gamma_j - 1} = \left(\frac{\theta_v}{\theta_j}\right)^{-b}, \tag{15}$$

where Γ_j is the average Lorentz factor of material in the jet core.

The adopted value for a and b must be representative of the structure very close to the core jet ($\theta_j < \theta_v \lesssim 2\theta_j$), which can differ from the shallow slopes $a, b \sim 5.5, 3.5$ inferred for the kinetic energy content of the jet in GW170817 (Ghirlanda et al. 2019) and that describe the structure up to very large angles $\theta_v \gtrsim 5\theta_j$. Furthermore, these structure indices inferred from afterglow data need not be equal to the indices we introduced in equations (14) and (15), if the kinetic-to-prompt conversion efficiency is not constant within the jet. As described in our plateau study (Beniamini et al. 2020a), steep slopes $a = 8$ and $b \gg 1$ are required to produce a plateau behaviour (their fig. 1). Thus, we adopt in our numerical exploration $a = 8$, $b = 6$, and $\Gamma_j = 250$. This structure of the Lorentz factor plays no role in the analytical developments of this section. However it is a prerequisite of our model that the misaligned observer's prompt emission is dominated by emission from the material on their line of sight. As shown in Appendix B, this condition is fulfilled with these adopted values for the slightly misaligned regime of viewing angles we are considering ($\theta_j < \theta_v < 2\theta_j$).

Of course, for both energy and Lorentz factor, the true structure may be more complicated and these prescriptions only concern the average material properties at a given angle from the jet axis. In particular, one should keep in mind that we suggest that flares are produced by individual shells in the core, for which properties such as Lorentz factor can deviate significantly from the average. A highly relativistic core with $\Gamma_j = 250$ that allows for ordinary prompt emission for an aligned observer does not exclude that some material in the core has a much lower Lorentz factor, as preferred for the production of flares in our model (see Section 2). The fact that most GRB afterglows with flares only present one or two flares (Chincarini et al. 2010) – in contrast with the many pulses in a GRB prompt phase – also suggests that only some of the shells produce flares.

With the assumed jet energy structure equation (14), we obtain the following flare–ESD contrast at the time of the flare

$$\frac{L_{flare}}{L_{ESD}^{LOS}(t_{flare})} = f_{geo}\mathcal{S}^{-3}\frac{\tau^c}{\tau^{LOS}}\left(\frac{\theta_v}{\theta_j}\right)^a\left(\frac{t_{flare}}{\tau^c}\right)^3 \tag{16}$$

$$= f_{geo}\frac{\tau^c}{\tau^{LOS}}\left(\frac{\theta_v}{\theta_j}\right)^a\left(\frac{\mathcal{S}^{-1}t_{flare}}{\tau^c}\right)^3 \tag{17}$$

$$= f_{geo}\left(\frac{\tau^c}{\tau^{LOS}}\right)^2\left(\frac{\theta_v}{\theta_j}\right)^a\left(1 - \frac{t_{ej}}{t_{flare}}\right)^{-3}, \tag{18}$$

where we used $t_{flare} = \mathcal{S}\tau^c + t_{ej}$, i.e. $\mathcal{S}^{-1}t_{flare} = \tau^c\left(1 - t_{ej}/t_{flare}\right)^{-1}$.

First, it appears from equation (18) that a steeper structure favours the appearance of flares during the ESD. While this is true – it is simply that the ESD is dimmer for a steep structure, one must bear in mind that a steeper structure also suppresses the misaligned prompt emission, and thus hinders the parent GRB detection and flare observation altogether. A finer calculation considering the likeliness of observing the GRB taking the structure into account is thus called for, see Section 5.3 for details.

Second, we find that, for fixed τ's, a later ejection time favours bright flares. This was already noted in our discussion of Fig. 2 and equation (11): One always has $t_{flare} = \mathcal{S}\tau^c + t_{ej} \geq t_{ej}$ and a larger t_{ej} favours a small $\mathcal{S}$ and thus a smaller suppression of flare flux by the $\mathcal{S}^{-3}$ factor. In the context of a flare during the ESD phase, it is clear that attributing a larger portion of the flare arrival time to ejection delay simply allows the flare to appear when the ESD flux is lower, as is clear in equation (18).

Finally, the appearance of the flare is dependent on the ratio of the core and line-of-sight angular time-scales τ. The τs depend on the material's Lorentz factor and dissipation radii. In general, the ratio of τ between two different lines of sight is not known, and is strongly model dependent.

On the one hand, if generally $\tau^c < \tau^{LOS}$, a bright flare can only be obtained if the ejection time is close to the time of the flare, as we just discussed. This may appear as a strong condition and leads us to discuss once more whether the central engine activity must be the same in all directions. If this activity is more prolonged in the core as compared to misaligned directions, there is more flexibility for large t_{ej} and bright flares, even if $\tau^c < \tau^{LOS}$. On the other hand, if $\tau^c > \tau^{LOS}$ is possible, this constraint does not apply and a flare can be produced without imposing a strict constraint on the duration of the central jet activity. Once again, we find that the central engine activity duration is a sensitive point in our model, and we will discuss more in Section 5.2.

It should finally be noted that the above considerations are valid for the average pulse. Depending on the actual pulse shape and luminosity, equation (18) may under or overestimate the actual contrast by a factor of a few. Also, spectral effects, not included in this section, could affect the flare visibility but are expected to remain moderate since both the ESD and the flares shine primarily in the X-rays.

3.2 A natural mechanism for narrow flare production

As mentioned in the introduction, a salient property of X-ray flares is their temporal aspect ratio w/t_{flare}, where w is the flare width. Flare widths have been measured in a number of ways: adopting for w the width of a Gaussian fit to the flare light curve (e.g. Chincarini et al. 2007); a smoothly broken power-law profile (Yi et al. 2016); or using the *Norris profile* (Norris et al. 2005), in which the width w is naturally given by the time span between the two points before and after the peak when the flare flux is a factor of e below the peak flux (Chincarini et al. 2010; Bernardini et al. 2011). Using this last definition, the aspect ratio w/t_{flare} – which is independent of redshift – is found to be tightly distributed around $w/t_{flare} = 0.23$ with a standard deviation of 0.14 (Chincarini et al. 2010). In other words, flares' aspect ratios are small and nearly constant in the population. Any flare model must reproduce this fact.

---

[1]Note that the below definition pertains to the emitted energy in the prompt phase, in contrast to other definitions of jet structure which concern the initial kinetic energy in the jet.

We will now exhibit a natural mechanism built into our misaligned observer picture for flares by which the brighter flares tend to be thin. While we will later estimate w/t_{flare} for actual flare light curves in our model (Section 3.3), we will temporarily focus on the temporal slope of flares in their post-peak phase, denoted by $\sigma = |\text{dlog } L(t)/\text{dlog } t|$. Analytically, we have a better grasp on σ than on w, and for any given profile, larger σ is related to smaller w, allowing us to focus on σ for now.

We introduce the time-dependent stretch factor $S(t)$ (note the different font) which is simply the $\mathcal{S}$-factor of the core material the radiation of which is received at time t by the misaligned observer. The calculation leading to equation (5) is still valid when applied only to a strip of the shell from which the observer receives radiation at time t. The f_{geo} term changes with time, as the strip will progressively scan the geometry of the shell as seen from the observer's stand point. However, for a generic shell shape, f_{geo} will not change significantly, and we will thus have $L_{\text{flare}}(t) \sim f_{\text{geo}} L_{\text{on}} S(t)^{-3}$. Finally, writing $S(t) = (t - t_{\text{ej}})/\tau^c$ as per equation (9), we arrive at the following approximate flare time behaviour

$$L_{\text{flare}}(t) = f_{\text{geo}} L_{\text{on}} \left(\frac{t - t_{\text{ej}}}{\tau^c} \right)^{-3}. \tag{19}$$

Taking the logarithmic derivative of equation (19) at $t = t_{\text{flare}}$, we find that the initial post-peak decay index for the flare is

$$\sigma = \left| -3 - \frac{3 t_{\text{ej}}}{\mathcal{S} \tau^c} \right| \tag{20}$$

$$= \frac{3}{1 - \frac{t_{\text{ej}}}{t_{\text{flare}}}}. \tag{21}$$

Therefore, coming back to equation (18), we find that the flare–ESD contrast is linked to the initial decay index by

$$\frac{L_{\text{flare}}}{L_{\text{ESD}}^{\text{LOS}}(t_{\text{flare}})} = f_{\text{geo}} \left(\frac{\tau^c}{\tau^{\text{LOS}}} \right)^2 \left(\frac{\theta_v}{\theta_j} \right)^a \left(\frac{\sigma}{3} \right)^3. \tag{22}$$

This last equation shows that, under our interpretation of flares, and during the ESD, brighter flares – or simply those that rise above the continuum – tend to decay faster, and therefore be thinner, explaining the low values observed for w/t_{flare}. This mechanism of selection of thin flares is particularly effective for small τ^c/τ^{LOS}. In this case, indeed, flares require even larger σ (i.e. must be even thinner) to appear above the continuum. However, as mentioned in Section 3.1, the relative values of τ^c and τ^{LOS} in GRB jets is uncertain.

Moreover, the width of flares in our picture is $w \propto \mathcal{S} \Delta t_{\text{on}} \sim \mathcal{S} \tau^c$, because the prompt pulse duration transforms like the photon arrival times to the off-axis line of sight. It therefore follows that the aspect ratio will be

$$\frac{w}{t_{\text{flare}}} = \eta \frac{\mathcal{S} \tau^c}{\mathcal{S} \tau^c + t_{\text{ej}}}, \tag{23}$$

where η is a proportionality constant depending on how exactly the width is measured.

As long as the arrival time is dominated by the angular effect ($\mathcal{S} \tau^c$) and not by the ejection time, i.e. $t_{\text{ej}}/\mathcal{S} \tau^c \ll 1$, all flares should have the same aspect ratio, in agreement with the observed tight distribution. Dispersion is expected due to flares with late ejection times and smaller values of $\mathcal{S}$, i.e. to flares where $t_{\text{ej}}/\mathcal{S} \tau^c \lesssim 1$. The explanation of the small scatter in aspect ratio is thus tightly linked to the expected range in t_{ej} and $\mathcal{S}$, which we discuss in Section 5.2.

3.3 X-ray flares from prompt dissipation in the jet core

For more concreteness we will now prescribe the emission physics for the core shell and study the production of flares for the misaligned observer. We assume $\theta_j = 0.1$ for the opening angle of the core jet and place the observer at $\theta_v = 0.13$ rad, at the same location as in our previous computation of plateau emission (Beniamini et al. 2020a; Fig. 1). The spherical shells responsible for the flares are in the core jet and, geometrically, cover the entire core of the jet up to a latitude of $\theta = \theta_j = 0.1$ rad. We suppose they instantaneously radiate a source-frame isotropic equivalent energy of $E_{\text{iso}}^c = 10^{53}$ erg while moving at a Lorentz factor of $\Gamma^c = 100$ in the radial direction. This set-up corresponds to $\mathcal{S} = 10$. We assign the shells with different values of t_{ej} and τ^c. We determine the light curves of the emission from these shells by integrating on equal-arrival-time strips on the shells corresponding to the angular exploration from the near to the far edge of the shell. All the light curves that we show are integrated in the *Swift*/XRT band $\mathcal{B}_{\text{XRT}} = [0.3-10]$ keV (Gehrels et al. 2004; Burrows et al. 2005a).

In order to produce the ESD radiation, we consider a shell along the observer's line of sight geometrically covering the entire jet from the line of sight to the core, which dissipates an energy $E_{\text{iso}}^{\text{LOS}} = 1.2 \times 10^{52}$ erg, consistent with the E_{iso}^c above and a steep power-law jet energy structure (equation 14) with $a = 8$ immediately outside the core as mentioned previously. The shell producing the ESD is supposed to be the last flashing shell, to which we assign an ejection time of $t_{\text{ej}}^{\text{LOS}} = 30$ s and a pulse duration of $\tau^{\text{LOS}} = 2$ s. The line-of-sight shell has a Lorentz factor $\Gamma^{\text{LOS}} = 50$, corresponding to a power-law structure (equation 15) with a core average Lorentz factor $\Gamma_j = 250$ and a slope $b = 6$. We recall that this is the same jet structure as assumed in our previous work on plateaus (Beniamini et al. 2020a).

For all core shells, we adopt a comoving emission spectrum of broken power-law shape, with low- and high-energy slopes $\alpha + 1 = -0.1$ and $\beta + 1 = -1.2$ (average 'F_ν' slopes found in the GRB prompt phase; Poolakkil et al. 2021) and a shell-frame peak energy $E_p' = 1$ keV, corresponding to the average observed E_p of ~ 200 keV for an aligned observer and a Lorentz factor of 100. With their $\mathcal{S} = 10$, all the flares would thus have a peak energy of 20 keV.

In Fig. 3, one can find the resulting flare light curves for the core shells (coloured lines) and the line-of-sight shell (ESD, black line). For completeness, we also added the level of a plateau predicted by our misaligned plateau model (Beniamini et al. 2020a; equation 10), with parameters exactly as in their fig. 1, with an external density $n = 1$ cm^{-3} for the uniform circum-burst medium hypothesis and a wind parameter $A_* = 0.1$ for the wind hypothesis. With these values adopted, it is as if the core shell dissipated $\eta_\gamma \sim 10$ per cent of the initial available kinetic energy in the core to produce a flare, and the rest served to produce the afterglow plateau. While this seems large at face value, it is consistent with typical prompt emission efficiencies estimated from data (Beniamini et al. 2015). Furthermore, the actual energy involved may be much less than this amount, if the flare producing shell is narrower than the jet's core (see discussion in Section 4).

The parameters for the flares can be found in Table 1: We provide the t_{ej} and τ and report the flare time t_{flare}, the aspect ratio w/t_{flare} as determined with the two points in time corresponding to a flux smaller by a factor of e than the peak flux, the contrast $L_{\text{flare}}/L_{\text{ESD}}$, and finally, the peak luminosity in the *Swift*/BAT band ([15–350] keV, Gehrels et al. 2004; Barthelmy et al. 2005) that an observer would detect if they were aligned with the shell.

First, Fig. 3 shows that our model is capable of producing thin flares with occurrence times and luminosities consistent with typically
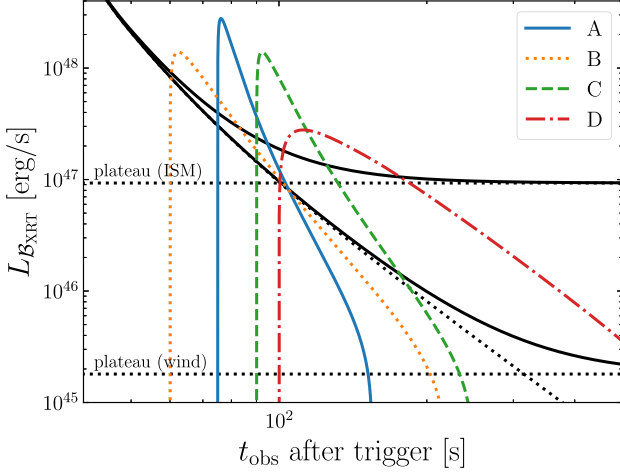
Figure 3. X-ray luminosity expected from flashing core shells, viewed as flares by an off-core observer. Coloured lines: Luminosity from the core shells. Black lines: ESD signal, produced by material along the line of sight, along with typical plateau levels predicted by our misaligned observer plateau model. The parameters for the shells are reported in Table 1.

observed flares. We note that the aspect ratio of the flare in a more realistic approach would be slightly affected by the intrinsic duration of the dissipation process in the core jet, leading to a less steep rise, and by the angular size of the shell, as briefly discussed in Section 4 where we fit our model to *Swift*/XRT data of two flares. The small values of the aspect ratio are one of the most puzzling feature of observed flares. Therefore the low values of w/t_{flare} listed in Table 1 are especially encouraging. Second, Table 1 shows that, as seen on axis, these flares would produce peak luminosities of $\sim 10^{52}$ erg/s in the BAT band, therefore their properties are consistent with being able to reproduce the gamma-ray emission.

Furthermore, we find that, at given τ^c (e.g. flares B and C), thinner flares tend to have a larger flare–ESD luminosity ratio, as discussed in Section 3.2, though we find the effect to be more pronounced than expected. Similarly, for a fixed flare–ESD luminosity contrast (e.g. flares A and C), it appears that a larger τ^c results in a flare with a larger aspect ratio, in line with equation (22) (recall that a larger σ is equivalent to a thinner flare). We also find that the larger t_{ej}/t_{flare} (still ≤ 1 of course), the steeper the post-decay phase and the thinner the flare, as expected from equation (21). The extreme case here being flare A, with $t_{ej}/t_{flare} \sim 92$ per cent (because of its small τ^c) resulting in an extremely steep initial decay, with $\sigma \gtrsim 6$. Finally, we note that obtaining later flares requires a larger τ^c (e.g. flare D). This, however, produces wider and dimmer flares, as

predicted in Section 2. In conclusion, this application of the model with instantaneous dissipation in the shells is consistent with the analytical results and observed trends in X-ray flares.

4 FLARE MORPHOLOGY

Beyond the width of flares, it appeared in the first catalogue (Chincarini et al. 2007) that flares presented a variety of morphologies, with rising and decaying phases being fast or slow, i.e. exponential or power-law profiles. In terms of actual rise and decay times, virtually all flares decay in a longer time than they rise (Chincarini et al. 2010). However, the temporal profiles near flare peak define either very localized peaks in the case of fast rise and decay, or rounder flares in the case of slow rise and decay. Here, we will further assess the capabilities of our misaligned-observer interpretation of flares by studying which flare morphology it is able to capture.

We seek to fit our model to GRB060719 (at $z = 1.532$) and GRB100816 (at $z = 0.8034$), which we choose because they feature flares peaking at $t_{flare,RF} \sim 90$ s in the source frame in both cases. The first presents a slow rise followed by a fast peak, while the second has a rounder peak and a slower decline, thus representing the variety in flare morphology. In Fig. 4, one can find the XRT data points for these two flares corrected for redshift. We applied the redshift correction by determining the source-frame time with $t_{RF} = t_{obs}/(1 + z)$ and source-frame luminosity in the redshifted band with

$$L_{(1+z) \times [\nu_{1,obs}, \nu_{2,obs}]} = \frac{1}{1+z} 4\pi D_L(z)^2 F_{[\nu_{1,obs}, \nu_{2,obs}]}, \quad (24)$$

where $F_{[\nu_{1,obs}, \nu_{2,obs}]}$ is the observer-frame flux and, in our case, $[\nu_{1,obs}, \nu_{2,obs}] = \mathcal{B}_{XRT} = [0.3-10]$ keV. We determined the luminosity distance $D_L(z)$ using a generic flat world model with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. The observer-frame data were retrieved from the *Swift*/XRT online archive (Evans et al. 2007, 2009).

We proceed by fitting the source-frame XRT light curve by the sum of a broken power-law light curve representing the continuum component and our flare model from core shells representing the excess calculated as in Section 3.3. For a reminder, the off-axis shell model has the following parameters: isotropic-equivalent dissipated energy E_{iso}, shell Lorentz factor Γ, angular distance between the observer and the shell's close edge $\Delta\theta = \theta_v - \theta_j$, shell ejection time t_{ej}, and decay time-scale τ. The emitted spectrum adopted is the same as in Section 3.3.

We seek the best-fitting model as determined by minimizing a χ^2 statistic under the same parameter constraints as in Section 2: E_{iso} in the range $10^{52.3 \pm 1}$ erg, τ in the 0.1–5 s range, $\mathcal{S}$ must be $\gtrsim$ 10. Concerning t_{ej}, we consider two different conditions: Either we

Table 1. Parameters for all flares with light curves represented in Fig. 3. t_{ej}: ejection time of the core shell responsible for the flare; τ^c: angular time-scale $R_e/2\Gamma^2 c$ of the shell, where Γ is the shell's Lorentz factor and R_e is the dissipation radius; t_{flare}: peak time of the flare; w: width of the flare, measured as the interval between the two times when the flux is $1/e$ times the peak flux; L^{flare}/L^{ESD}: ratio of the peak flux of the flare to the ESD flux at flare peak; $L_{p,BAT,on}^{prompt}$: peak pulse luminosity that an observer aligned with the shell would detect in the *Swift*/BAT band. All shells release the same isotropic equivalent energy $E_{iso}^c = 10^{53}$ erg; all flares have $\mathcal{S} = 10$. Note that our choice of gamma-ray band (*Swift*/BAT) and of emitted spectrum results in the approximate relation $L_{p,BAT,on}^{prompt} \sim 0.5 E_{iso}^c/\tau^c$.

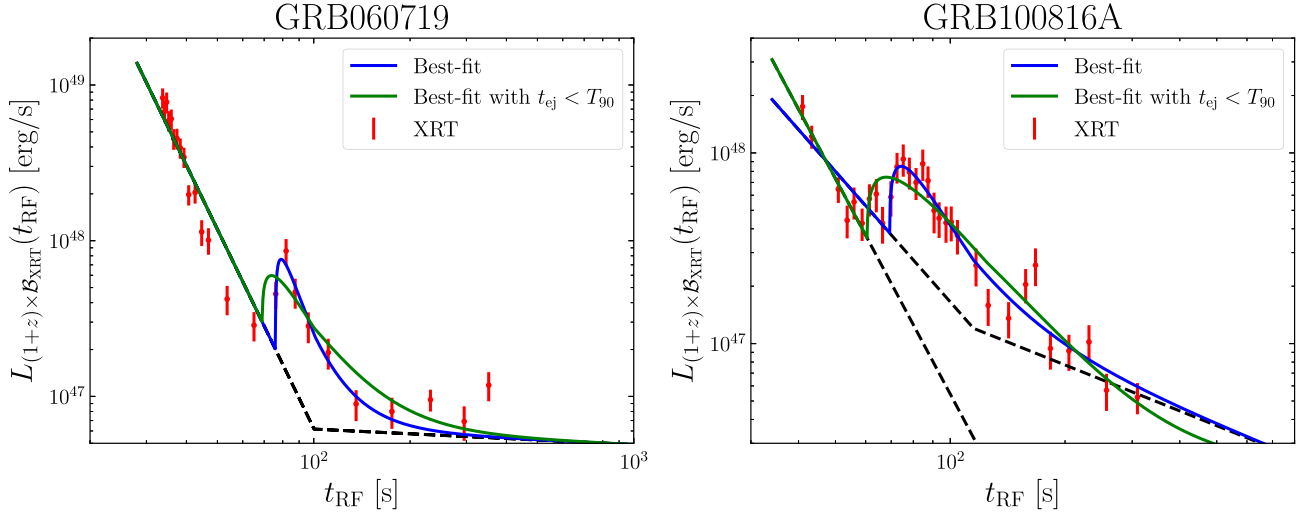| #	t_{ej} (s)	τ^c (s)	t_{flare} (s)	w/t_{flare}	L^{flare}/L^{ESD}	$L_{p,BAT,on}^{prompt}$ (erg s^{-1})
A	70	0.5	76	0.10	9.2	1.0×10^{53}
B	50	1	62	0.25	1.7	5.0×10^{52}
C	80	1	92	0.17	9.7	5.0×10^{52}
D	50	5	112	0.69	2.9	1.0×10^{52}

Figure 4. Joint continuum and excess fitting to source-frame XRT data for two flares with different morphologies, using our misaligned core shell dissipation model. Best fits are shown both constraining the shell ejection time to $\leq T_{90,RF}$ (green) or not (blue). The dashed black line shows the best-fitting continuum, which differs between the free-t_{ej} and constrained-t_{ej} fits only for GRB100816A (right). See Table 2 for the corresponding best-fitting parameters.

Table 2. Best-fitting source-frame parameters for XRT data of flares in GRB060719 and GRB100816A with our misaligned-observer interpretation of flares. We present parameters both constraining the ejection time to $t_{ej} \leq T_{90,RF}$ or leaving it unconstrained. Notations for flare parameters are the same as in Table 1.

GRB	060719 ($z = 1.532$, $T_{90,RF} = 26.4 \pm 4.5$)		100816A ($z = 0.8034$, $T_{90,RF} = 1.6 \pm 0.3$)	
	t_{ej} free	$t_{ej} < T_{90,RF}$	t_{ej} free	$t_{ej} < T_{90,RF}$
E_{iso} (erg)	6.9×10^{52}	1.0×10^{53}	3.8×10^{52}	1.8×10^{53}
Γ	119	87	78	90
$\Delta\theta$ (rad)	0.033	0.048	0.035	0.041
t_{ej} (s)	59	31	42	1.8
τ (s)	1.0	2.1	3.3	4.0
$\mathcal{S}$	17	18	8.6	15

let it vary freely up to $t_{flare,RF}$ or we bound it by $T_{90,RF}$, the source-frame GRB duration. Coming back to our discussion on central engine activity, the latter constrained-t_{ej} condition hypothesizes that this activity's duration is the same in all directions to the central engine, therefore having measured it through T_{90} along the line of sight constrains it in the core; the former free-t_{ej} condition does not make this hypothesis. In Fig. 4 we show the best fits under both these conditions; best-fitting source-frame parameters can be found in Table 2.

First, it seems that satisfactory fits can be found to the light curve with reasonable parameter values: in both flares, the energies remain within the allowed region, $\Gamma \sim 100$ as anticipated in Section 2 and the $\mathcal{S}$-factors are $\gtrsim 10$, within the slightly misaligned regime. As a result, $t_{ej} < 0.5 t_{flare}$, meaning both the delayed ejection and the light traveltime effects are at play.

In these fits as in our examples of Section 3.3, we supposed that the shells geometrically covered all the core, such that the real jet dissipated energy is

$$E_{real}^c \sim \frac{\theta_j^2}{2} E_{iso}^c. \tag{25}$$

By changing the sizes of the shells (i.e. their angular diameter), we found that the observable section of the flare light curves do not change. The only signature of the size of the shell is the sharp cutoff in flux produced by the far edge of the shell, and which can be noted

for flares A, B, and C in Fig. 3. In all cases however, this occurs at a flux level much lower than the underlying continuum and is therefore not visible to observers. In principle, the angular size of the shells are not constrained by the physics in the jet, apart from causality arguments implying their angular size cannot be smaller than $1/\Gamma$. As the shell size has little influence on the resulting flare, it allows some liberty to decrease the actual energy budget E_{real}^c of the model.

Second, it is clear that the hypothesis of instantaneous dissipation in the comoving frame leads to a sharp rise in the flares, and therefore allows our model to better fit fast-rising morphologies such as in GRB100816A rather than slow-rising ones like in GRB060719.

While instantaneous dissipation was practical for the analytical calculations above, it is not a requirement of our model. An intrinsic duration of dissipation will mainly affect the rise of the flare. However the properties of the declining phase – notably, the initial decay slope discussed in Section 3.2 – should remain, as this phase is dominated by the angular exploration of the shell to latitudes further from the observer (this is equivalent to the situation in rise-decay ratios for prompt pulses, see section 4.4 of Beniamini & Granot 2016). In particular, the transformation of the duration of a pulse to the duration of the flare (equation 9) is still valid, when restricted to the declining phase of the pulse and flare.

Similarly, a shell shape different than circular or a unequally bright shell would change the profile of the rising phase. The circular shape we adopted is not particularly physically motivated, and the uniform

shell brightness allowed simplifying some derivations; different prescriptions would not change the main features of our model. We thus conclude that an intrinsic dissipation duration and other shell shapes would allow to better capture slow-rising flares such as in GRB060719. However, the declining phase in both examples is well captured by the model, whether steep (GRB060719) or slow (GRB100816A).

Third, it is obvious that letting the ejection time run free makes for much better fits. This is true for GRB060719, where the $t_{\rm ej} \leq T_{\rm 90,RF}$ restriction seems to not allow the flare to peak at the right time. A similar issue, though less pronounced, occurs for GRB100816A. We did not statistically compare the goodness of fit of the free-$t_{\rm ej}$ and restricted-$t_{\rm ej}$ models. Nonetheless these two examples show that this hypothesis plays an important role in our model for flares, and we discuss it in more detail in Section 5.2.

5 DISCUSSION

5.1 Summary

Motivated by modelling for plateaus in GRB afterglows in the physical setting of a slightly misaligned observer to a structured jet (Eichler & Granot 2006; Beniamini et al. 2020a), we have presented a new model to interpret flares in the same set-up. We suggest that flares can be produced by prompt dissipation in the jet's core that appears in the X-ray band rather than the gamma-rays because the core is less boosted to a misaligned observer than to an on-axis observer. In our picture, the delay in the flare observation with respect to the prompt emission is a combination of a purely geometrical effect, linked to the photon traveltime from the core to the misaligned line of sight, and an intrinsic effect linked to the finite duration of the central engine activity, resulting in delays in ejection of different shells.

Writing down the transformations of photon arrival time, spectrum, and luminosity from an aligned to a misaligned line of sight, we found that the typical X-ray flare could indeed be explained as deboosted core jet prompt emission (Section 2). In doing so, we outlined the typical properties required for the core shells responsible for the flares and the expected properties for the resulting flares: Those shells have rather low Lorentz factors $\Gamma \sim 100$, carry energies typically on the higher end of those dissipated in single pulses of GRBs $E_{\rm iso} \gtrsim 10^{52}$ erg. The flares thus produced naturally occur early in the afterglow ($t_{\rm flare} \leq 1000$ s), when the majority of flares are observed in XRT light curves.

In the early afterglow phase, the continuum is dominated by the ESD. We therefore analytically studied the conditions for appearance of flares during the ESD, assuming it was produced by high-latitude emission from the last flashing shell on the misaligned observer's line of sight (Section 3). We found that flare visibility is favoured by late shell ejection times $t_{\rm ej}$ and long shell decay time-scales $\tau_{\rm c}$. We also demonstrated a mechanism present in our model by which brighter flares – or simply those that are able to appear above the continuum – tend to be narrower, echoing the observation that most flares have aspect ratios $w/t_{\rm flare} \lesssim 0.5$. Furthermore, synthetic light curves confirmed these trends between brightness and aspect ratio, the prominent role of the ejection time in flare visibility and overall, proved that typical flare widths are reproduced by our model (Section 3.3).

Finally, we made fits of our model to two actual flares observed in XRT afterglows (Section 4). We chose these two examples so as to represent two different morphologies found in flares: slow-rise-fast-decay (GRB060719) and fast-rise-slow-decay (GRB100816A).

We found satisfactory fits with reasonable parameter values and a set-up within the slightly misaligned regime, as measured by $\mathcal{S} \sim 1 + |\Delta\theta\Gamma|^2 \gtrsim 10$. The model would better capture the slow rises assuming an intrinsic duration of the shell dissipation – our analytical work considered it instantaneous for simplicity.

These fits confirmed the role of the ejection time in defining the flare arrival time and the trade-off between shell energy and ejection time anticipated in Section 2. Indeed, flare arrival time increases with both $\mathcal{S}$ (more misalignment) and $t_{\rm ej}$ (later ejection), flare luminosity however drastically decreases with $\mathcal{S}$. Thus, for a given flare, increasing $t_{\rm ej}$ allows to somewhat decrease $\mathcal{S}$ and the shell energy.

5.2 Admissible prompt dissipation mechanisms

In principle, our explanation of flares and our analytical results are compatible with any prompt dissipation mechanism, as long as the declining phases of gamma-ray pulses are due to the angular exploration of the shell, i.e. high-latitude emission. This will be the case for prompt mechanisms with negligible dissipation duration compared to the angular time-scale, such as internal shocks (Daigne & Mochkovitch 1998, 2000) or macroscopic magnetic reconnection events at large distances from the central engine (Lyutikov & Blandford 2003; Kumar & Narayan 2009; Kumar & Crumley 2015; Barniol Duran, Leng & Giannios 2016; Beniamini & Granot 2016; Beniamini, Barniol Duran & Giannios 2018). It could also be the case for models with smaller dissipation radii, such as dissipative photospheric models (e.g. Rees & Mészáros 2005; Beloborodov & Mészáros 2017), if the central engine turns off sufficiently rapidly to mimic a steep decline as in high-latitude emission. All the other prompt-related parameters of the model such as energy and shell Lorentz factor are generic. Because of this last point, the occurrence of the mechanism we propose is inevitable: Any misaligned observer to a GRB jet is exposed to detecting delayed and deboosted prompt emission from the jet core. Our fit to the flare in GRB100816A occurring 100 s after trigger with $t_{\rm ej} \sim 0$ (Fig. 4, right-hand panel) shows that even a null ejection delay can trigger the mechanism. There must therefore be some flares produced in this way. As already mentioned, these should be early flares because of the allowed parameter ranges. The statistical properties of the flares produced by our mechanism will be further discussed in Section 5.3 below.

One subtle point of our model is the admissible range for the ejection time of the core shells. As we mentioned, larger $t_{\rm ej}$ allows for brighter, thinner flares. In our picture, the shell ejection must naturally occur during the first and only episode of central engine activity. Having observed a GRB as a misaligned observer informs us about the central engine activity duration along one's line of sight which dominates the observed signal. This duration has no reason to be the same than that of the ejection activity in the jet core, where we posit the flare-producing shells lie. The example of GRB100816A (Fig. 4, right-hand panel) shows that a better fit is found by allowing $t_{\rm ej}$ to reach typical GRB T_{90} durations of 40 s which, however, are much longer than the parent GRB's source-frame T_{90} of 1.6 s. GRB060719 (Fig. 4, left-hand panel) provides a less drastic example.

These results suggest the question of whether the central engine activity can be shorter or longer lasting depends on the ejection direction. Such variation in the activity duration around the central engine is prescribed upstream, by the physical conditions near the compact source. Because of, e.g. the significant interaction of the incipient jet with material near the compact object, it could be that the

time interval during which relativistic material effectively emerges from the system depends on the latitude. One could speculate that the jet breaks out early and continues to eject matter through the core while the off-axis material experiences more interaction with the cocoon and thus relativistic material emergence is shorter lived. To our knowledge, the comparison of relativistic ejection of different lines of sight remain to be studied in numerical simulations, and more fits of our model to XRT data must be done to determine whether the requirement of $t_{ej} \geq T_{90}$ is a general feature.

Generally, a constant w/t_{flare} cannot be a natural consequence of a *composite model*, in which w and t_{flare} are set by unrelated causes. For example, in genuine late central engine activity (e.g. King et al. 2005; Dai et al. 2006; Perna et al. 2006), w and t_{flare} are determined respectively by the duration and onset time of the second engine episode. In this case, the tuning of w/t_{flare} to a same value of the aspect ratio from one system to another is not clear, especially seeing the diversity of durations in the first episode, as suggested by GRB durations. In models with changes in the reverse shock propagation medium (e.g. Hascoët et al. 2017), w and t_{flare} are determined by the size and the propagation time up to the external medium accident, tuning these also seems unnatural. A similar limitation affects pictures for flares including a stratified propagation medium for the reverse shock (Lamberts & Daigne 2018; Ayache et al. 2020) or those with a delayed prompt dissipation (Giannios 2006; Yu & Dai 2009; Troja et al. 2015). We also note that such models also fall short of explaining the steep rises of some flares (e.g. in GRB060719, Fig. 4, left-hand panel).

Conversely, a constant w/t_{flare} is a natural consequence of models with a single episode of central engine activity and a single episode of prompt dissipation, such as ours and the model of Beniamini & Kumar (2016). In these cases, the aspect ratio is determined by a single transformation of prompt emission, in our case by geometrical effects, i.e. photon traveltime. The fact that the distribution of rise-to-decay time ratios for flares closely follows that of GRB prompt pulses (Chincarini et al. 2010) further encourages such models. In our picture, we showed in Section 3.2 that a small scatter in flare aspect ratio is obtained by diversity in t_{ej} and $\mathcal{S}$: We expect the geometrical effect ($\mathcal{S}\tau^c$) to be dominant in shaping flare aspect ratio, thus producing a tight distribution, and the subdominant t_{ej} to introduce a small scatter.

Another remarkable property of our model is that the flares naturally appear in the X-rays, even if gamma-rays are produced for an aligned observer. It is not obvious why the central engine should shine in the X-rays and with an increasing time-scale of variability in other models. In fact, because of the misaligned nature of the observer in our picture, no simultaneous higher energy counterpart (e.g. gamma-rays) is expected. Indeed, it seems that any super-gamma-ray photons which would eventually appear as gamma-rays to the misaligned observer should be suppressed by optical depth to pair production, scattering on pairs and on electrons (Matsumoto, Nakar & Piran 2019a, b); the X-rays however are not affected. A thorough establishment of this prediction in our geometrical set-up remains to be done. We checked that no simultaneous BAT counterpart was present in the two examples treated above in Section 4. While we have mainly discussed flares in the X-ray band, they are also present in the optical, where they are found to be statistically similar to X-ray flares (Li et al. 2012; Yi et al. 2017) and to also occur mostly in the early afterglow (Swenson et al. 2013). This could indicate a common origin for optical and X-ray flares as well, as suggested here. There remains to explore the optical signatures of our picture for flares. Surely the fitting of multiwavelength data will further constrain the parameter space.

5.3 Observed and expected model consequences

There are rather strong conditions for successful flaring in our model. Mainly, these are large E_{iso} and small Γ in the shells responsible for the flares, as dictated by the transformation of luminosity from the aligned to misaligned lines of sight (equation 10). These conditions are further restricted by finer effects, for example the fact that later GRB prompt pulses tend to be less energetic than earlier ones (Ghirlanda, Nava & Ghisellini 2010), whereas our model favours the appearance of flares from shells ejected later rather than earlier. Therefore, we do not expect all the shells in a given GRB jet to produce visible flares, and we do not expect many flares in each GRB. Adding the condition of slightly misaligned line of sight to the GRB jet, we do not expect many GRBs with flares in general; the observation is that about one third of *Swift* GRBs exhibit noticeable flares (Yi et al. 2016). Discussing such statistics requires to consider many parameters: the allowed range for t_{ej}; the jet structure, which conditions both the GRB trigger for off-axis observers and the flare-to-ESD contrast (equation 18); and the Lorentz factors of the shells within the core. Such a statistical study would further root our model into flare observations. Moreover, as the physical set-up is the same for our two models, statistics of both flaring and plateau behaviour in GRB afterglow would shed further light on our models. Such joint flare–plateau activity has already been noted (e.g. GRB080129, Gao 2009; Greiner et al. 2009) and a recent study has quantified the proportion of GRBs with plateaus and flares to ∼50 per cent among GRBs with flares (Yi, Du & Liu 2022). In this subsample, strong correlations are found between the energy dissipated in the parent GRB prompt phases and in the flares. This observation is encouraging for our picture of plateaus and flares, and we leave a thorough statistical study to future work.

An interesting consequence of our model is the natural dichotomy between early and late flares, with the former class being statistically dominant (Swenson & Roming 2014). With t_{ej} limited to $\lesssim 100$ s (i.e. typical central engine activity), flares with $t_{flare} \geq 1000$ s are only possible in our picture with large $\mathcal{S}$. However with $L_{flare} \propto \mathcal{S}^{-3}E_{iso}$, we do not expect these flares to be visible. This fact remarkably echoes the observation of a dichotomy in temporal behaviour and $\Delta F/F$ distributions between early and late flares (Bernardini et al. 2011): A different origin for late flares is therefore reasonable.

It is also interesting that, for these early flares, Bernardini et al. (2011) found the relation $L_{flare} \propto t_{flare}^{-2.7\pm0.1}$ between flare arrival times and peak fluxes. Considering that $L_{flare} \sim \mathcal{S}^{-3}L_{on}$ and $t_{flare} = \mathcal{S}\tau^c + t_{ej}$, the correlation spanned by varying $\mathcal{S}$ for constant τ and shell energies should be $L_{flare} \propto t_{flare}^{-3}$ exactly, in the absence of shell ejection delays. These delays allow however for later flares with the same luminosity, such that the slope is in fact slightly shallower than -3, as found. Outlying flares of such luminosity-peak time correlations that are both late and bright (such as the 'giant flare' GRB050502B, Falcone et al. 2006, or the late flares in GRB121027A, Peng et al. 2013) are however out of reach of our model's parameter space. Finally, for GRBs with many flares, we expect the later ones to have larger $\mathcal{S}$'s generally, considering the other parameters fixed. We thus expect them to be dimmer and softer; these trends are indeed found in the few GRBs with more than one flare (Chincarini et al. 2010).

If our picture for flares is correct, there are further consequences that we could check in the population. First, while the question of dependence of T_{CE} on the line of sight is delicate, we should generally expect that, if T_{90} is large in a given GRB, then it should also be large on other lines of sight to the same GRB. As large ejection times favour flares, we expect long T_{90} to be correlated with

flaring activity. A preliminary inspection of the *Swift* archive seems to confirm this trend, though a thorough statistical study is called for. Second, we describe early flares and plateaus as the consequence of misaligned lines of sight, thus we expect the jet structure to introduce a trend between flaring activity and dim prompt emission. Third, if there is a qualitative or quantitative difference in the prompt dissipation mechanism between the core and the lateral structure, it should appear when comparing prompt emission of GRBs with and without flaring or plateau activity. Indeed in our picture we expect the prompt emitting regions in these two cases to be located in different parts of the jet structure in which, at least quantitatively, the physical conditions and thus prompt dissipation mechanism could be different. With this last point, we understand that flares and plateaus could provide key insight into GRB physics, the structure of GRB jets and the difference between long and short GRBs in this respect.

Generally, while the present publication is only a first proposal and limited exploration of this mechanism for flares, the many consequences of this model on the population of flares that we have just listed makes it testable and provides motivation to deepen the statistical study of GRB afterglow flares. In confronting our model's predictions with flare statistics and possibly seeking to constrain the jet structure, we note three main points of caution. First, we expect some – yet unknown – degree of diversity in GRB jet structures and core jet opening angle. Second, one must carefully handle selection effects in flare samples linked to incomplete redshift data or X-ray coverage of the early afterglow of GRBs. Third, one must account for detection biases intrinsic to our model: In our picture, the detection of flares is subject to the detection of the prompt emission from the off-axis line-of-sight material, which is itself linked to the energy and Lorentz-factor structures of the jet.

5.4 On the nature of flares and the flare sample

Our model posits a common origin for prompt emission and early X-ray flares, and is supported by some similarities between these phenomena. Consequently, it poses the question of the definition of flares and sample contamination in X-ray flares. All X-ray flare samples we mentioned in this paper (Chincarini et al. 2007, 2010; Falcone et al. 2007; Bernardini et al. 2011; Margutti et al. 2011; Swenson & Roming 2014; Yi et al. 2016) are selected only after visual inspection of the XRT light curves, in search for excess flux over a continuum. While this is justified for late flares (such as the late flare sample of Bernardini et al. 2011), the sample of early flares thus selected must be contaminated by prompt emission as well. Generally, a more physical definition of flares is warranted to better isolate this activity and thus define what exactly models should seek to reproduce. For example, in our picture, it seems that flares should be characterized by an absence of a counterpart in the higher energy bands such as the BAT and a significantly larger hardness ratio than the underlying continuum emission. This definition excludes, for example, the sample collected in Peng et al. (2014), the flares in GRB050820A (Cenko et al. 2006; $t_{flare} \sim 400$ s) and GRB110801A (De Pasquale et al. 2011; $t_{flare} \sim 200$ s), and the first two of the five flares in GRB060714 (Krimm et al. 2007; $t_{flare} = 80$–100 s), which show a simultaneous BAT excess.

Furthermore, we note that positing a common origin for GRB prompt emission and X-ray flares further adds to the discussion of the possibility that long GRBs, low-luminosity GRBs, X-ray flashes, etc. could be the manifestations of the same system viewed from different orientations (e.g. Granot, Ramirez-Ruiz & Perna 2005; Ramirez-Ruiz et al. 2005; Granot & Ramirez-Ruiz 2010). That fact that the low-luminosity GRB031203 and GRB171205A present some phases of flat X-ray flux (respectively, Ramirez-Ruiz et al. 2005; D'Elia

et al. 2018) supports the idea that they could be misaligned events, in light of our plateau model. In light of the present flare model, the existence of X-ray flashes with contemporaneous X-ray and optical flares (such as XRF071031, Krühler et al. 2009) also supports this view. The investigation of X-ray flashes as misaligned core prompt dissipation with a model as presented here could shed more light on the nature of these events. In any case, both a statistical study and light curve fits should be carried out.

Finally, we note that the present article and other recent afterglow modelling endeavours (e.g. Ascenzi et al. 2020; Beniamini et al. 2020b; Oganesyan et al. 2020) illustrate the strong effects of observer–jet geometry in shaping GRB prompt and afterglow emission, even without specifying the underlying dissipation mechanisms at play. The combination of these geometrical effects with structured jets makes for fruitful modelling grounds and allow for fresh views on the GRB phenomenon.

6 CONCLUSION

In the same physical set-up as a model to explain the plateau features in GRB afterglows, we exhibited a novel interpretation for X-ray flares in GRB afterglows. It relies on slightly misaligned lines of sight to a structured relativistic jet, in which the core's prompt dissipation is deboosted to the X-ray band for the off-axis observer and appears during the afterglow – typically, during the ESD – because of both the light traveltime from the core to the observer and the intrinsic duration of the central engine ejection activity. From order-of-magnitude considerations to actual fits to *Swift*/XRT data, we showed that this model is capable of explaining typical flares. Further, we showed how this model favours flares with small and tightly distributed aspect ratios, a salient property of X-ray flares. Overall, though a thorough statistical study is called for, it appears that there are many trends found in GRB afterglows flares that our picture naturally produces.

DATA AVAILABILITY

The data and software underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Amati L., 2006, MNRAS, 372, 233
Ascenzi S., Oganesyan G., Salafia O. S., Branchesi M., Ghirlanda G., Dall'Osso S., 2020, A&A, 641, A61
Ayache E. H., van Eerten H. J., Daigne F., 2020, MNRAS, 495, 2979
Barniol Duran R., Leng M., Giannios D., 2016, MNRAS, 455, L6
Barthelmy S. D. et al., 2005, Space Sci. Rev., 120, 143
Beloborodov A. M., Mészáros P., 2017, Space Sci. Rev., 207, 87
Beniamini P., Granot J., 2016, MNRAS, 459, 3635
Beniamini P., Kumar P., 2016, MNRAS, 457, L108
Beniamini P., Nakar E., 2019, MNRAS, 482, 5430

Beniamini P., Nava L., Duran R. B., Piran T., 2015, MNRAS, 454, 1073
Beniamini P., Barniol Duran R., Giannios D., 2018, MNRAS, 476, 1785
Beniamini P., Duque R., Daigne F., Mochkovitch R., 2020a, MNRAS, 492, 2847
Beniamini P., Granot J., Gill R., 2020b, MNRAS, 493, 3521
Bernardini M. G., Margutti R., Chincarini G., Guidorzi C., Mao J., 2011, A&A, 526, A27
Burrows D. N. et al., 2005a, Space Sci. Rev., 120, 165
Burrows D. N. et al., 2005b, Science, 309, 1833
Cenko S. B. et al., 2006, ApJ, 652, 490
Chincarini G. et al., 2007, ApJ, 671, 1903
Chincarini G. et al., 2010, MNRAS, 406, 2113
D'Elia V. et al., 2018, A&A, 619, A66
Dai Z. G., Wang X. Y., Wu X. F., Zhang B., 2006, Science, 311, 1127
Daigne F., Mochkovitch R., 1998, MNRAS, 296, 275
Daigne F., Mochkovitch R., 2000, A&A, 358, 1157
De Pasquale M. et al., 2011, GCN Circ., 12228, 1
Eichler D., Granot J., 2006, ApJ, 641, L5
Evans P. A. et al., 2007, A&A, 469, 379
Evans P. A. et al., 2009, MNRAS, 397, 1177
Falcone A. D. et al., 2006, ApJ, 641, 1010
Falcone A. D. et al., 2007, ApJ, 671, 1921
Fan Y. Z., Wei D. M., 2005, MNRAS, 364, L42
Gao W.-H., 2009, ApJ, 697, 1044
Gehrels N. et al., 2004, ApJ, 611, 1005
Ghirlanda G., Nava L., Ghisellini G., 2010, A&A, 511, A43
Ghirlanda G. et al., 2019, Science, 363, 968
Giannios D., 2006, A&A, 455, L5
Granot J., Ramirez-Ruiz E., 2010, preprint (arXiv:1012.5101)
Granot J., Ramirez-Ruiz E., Perna R., 2005, ApJ, 630, 1003
Greiner J. et al., 2009, ApJ, 693, 1912
Guidorzi C., Dichiara S., Frontera F., Margutti R., Baldeschi A., Amati L., 2015, ApJ, 801, 57
Hakkila J., Horváth I., Hofesmann E., Lesage S., 2018, ApJ, 855, 101
Hascoët R., Beloborodov A. M., Daigne F., Mochkovitch R., 2017, MNRAS, 472, L94
King A., O'Brien P. T., Goad M. R., Osborne J., Olsson E., Page K., 2005, ApJ, 630, L113
Kobayashi S., Zhang B., Mészáros P., Burrows D., 2007, ApJ, 655, 391
Krimm H. A. et al., 2007, ApJ, 665, 554
Krühler T. et al., 2009, ApJ, 697, 758
Kumar P., Crumley P., 2015, MNRAS, 453, 1820
Kumar P., Narayan R., 2009, MNRAS, 395, 472
Lamb G. P., Kann D. A., Fernández J. J., Mandel I., Levan A. J., Tanvir N. R., 2021, MNRAS, 506, 4163
Lamberts A., Daigne F., 2018, MNRAS, 474, 2813
Lazzati D., Perna R., 2007, MNRAS, 375, L46
Li L. et al., 2012, ApJ, 758, 27
Liang E. W. et al., 2006, ApJ, 646, 351
Lyutikov M., Blandford R., 2003, preprint (arXiv:astro–ph/0312347)
Malesani D. et al., 2007, A&A, 473, 77
Margutti R., Guidorzi C., Chincarini G., Bernardini M. G., Genet F., Mao J., Pasotti F., 2010, MNRAS, 406, 2149
Margutti R., Bernardini G., Barniol Duran R., Guidorzi C., Shen R. F., Chincarini G., 2011, MNRAS, 410, 1064
Matsumoto T., Nakar E., Piran T., 2019a, MNRAS, 483, 1247
Matsumoto T., Nakar E., Piran T., 2019b, MNRAS, 486, 1563
Mooley K. P. et al., 2018, Nature, 561, 355
Norris J. P., Bonnell J. T., Kazanas D., Scargle J. D., Hakkila J., Giblin T. W., 2005, ApJ, 627, 324
Nousek J. A. et al., 2006, ApJ, 642, 389
Oganesyan G., Ascenzi S., Branchesi M., Salafia O. S., Dall'Osso S., Ghirlanda G., 2020, ApJ, 893, 88
Panaitescu A., 2008, MNRAS, 383, 1143
Peng F.-k., Hu Y.-D., Xi S.-Q., Wang X.-G., Lu R.-J., Liang E.-W., Zhang B., 2013, preprint (arXiv:1302.4876)
Peng F.-K., Liang E.-W., Wang X.-Y., Hou S.-J., Xi S.-Q., Lu R.-J., Zhang J., Zhang B., 2014, ApJ, 795, 155
Perna R., Armitage P. J., Zhang B., 2006, ApJ, 636, L29

Pescalli A., Ronchi M., Ghirlanda G., Ghisellini G., 2018, A&A, 615, A80
Poolakkil S. et al., 2021, ApJ, 913, 60
Ramirez-Ruiz E., Granot J., Kouveliotou C., Woosley S. E., Patel S. K., Mazzali P. A., 2005, ApJ, 625, L91
Rees M. J., Mészáros P., 2005, ApJ, 628, 847
Salafia O. S., Ghisellini G., Pescalli A., Ghirlanda G., Nappo F., 2015, MNRAS, 450, 3549
Sonbas E., MacLachlan G. A., Shenoy A., Dhuga K. S., Parke W. C., 2013, ApJ, 767, L28
Swenson C. A., Roming P. W. A., 2014, ApJ, 788, 30
Swenson C. A., Roming P. W. A., De Pasquale M., Oates S. R., 2013, ApJ, 774, 2
Takahashi K., Ioka K., 2021, MNRAS, 501, 5746
Troja E., Piro L., Vasileiou V., Omodei N., Burgess J. M., Cutini S., Connaughton V., McEnery J. E., 2015, ApJ, 803, 10
Von Kienlin A. et al., 2020, ApJ, 893, 46
Yi S.-X., Xi S.-Q., Yu H., Wang F. Y., Mu H.-J., Lü L.-Z., Liang E.-W., 2016, ApJS, 224, 20
Yi S.-X., Yu H., Wang F. Y., Dai Z.-G., 2017, ApJ, 844, 79
Yi S.-X., Du M., Liu T., 2022, ApJ, 924, 69
Yu Y. W., Dai Z. G., 2009, ApJ, 692, 133

APPENDIX A: TRANSFORMATION OF BOLOMETRIC LUMINOSITY FROM ALIGNED TO MISALIGNED LINES OF SIGHT

We derive the transformation of the observed luminosity from a flashing shell from a given line of sight to another.

We start from the definition for the spectral luminosity:

$$L_\nu(t) = \int \frac{\mathrm{d}L_\nu(t)}{\mathrm{d}\Omega}\mathrm{d}\Omega. \tag{A1}$$

Where $\frac{\mathrm{d}L_\nu(t)}{\mathrm{d}\Omega} = \frac{\mathrm{d}E}{\mathrm{d}t\mathrm{d}\nu\mathrm{d}\Omega}$ is the emitted energy per unit time, frequency, and emitting region solid angle and we identify observer time and emission time, disregarding light traveltime delays. Changing frames from the emitter's frame to the observer's frame, dE transforms as D and dt and dν transform as D^{-1}, where $D = \frac{1}{\Gamma(1-\beta\cos\theta)}$ is the Doppler factor.

Therefore, we have

$$L_\nu(t) = \int D^3 L'_{\nu'}(t)\mathrm{d}\Omega. \tag{A2}$$

Different lines of sight to the source correspond to different segments of the jet dominating the received emission, i.e. the solid angles with the largest Doppler factor among those which radiate. We denote by θ_0 the total angular size of the emitting region and $\delta\theta$ the angular distance between the line of sight and the edge of the emitting region. We therefore have the following regimes:

(i) An aligned line of sight ($\delta\theta = 0$): The observed emission is dominated by a ring with $\theta < 1/\Gamma$, thus d$\Omega \sim 2\pi \times 1/2\Gamma^2$, the Doppler factor is $D \sim 2\Gamma$ and

$$L_\nu^{\mathrm{on}} \sim 8\pi\,\Gamma L'_{\nu'}. \tag{A3}$$

(ii) A slightly misaligned line of sight ($1/\Gamma \ll \delta\theta \ll \theta_0$): The flux is dominated by regions within the emitting region that are the most boosted, with $\delta\theta < \theta < 2\delta\theta$. In addition, the emitting region is limited to a transverse angular size of $\Delta\phi \sim \pi$ of the emitting region which occupies nearly a half plane in the observer's field. Thus d$\Omega \sim 3\delta\theta^2\Delta\phi/2$, the Doppler factor is $D = \frac{1}{\Gamma(1-\beta\cos\delta\theta)} \sim 2\Gamma/(1 + \Gamma^2\delta\theta^2)$ and

$$L_\nu^{\mathrm{off}} \sim \frac{8\Gamma^3}{(1 + \Gamma^2\delta\theta^2)^3}\frac{3\Delta\phi\delta\theta^2}{2}L'_{\nu'} \sim \frac{\Delta\phi}{2\pi}S^{-2}L_\nu^{\mathrm{on}}, \tag{A4}$$

where we have used $S = \frac{1-\beta\cos\delta\theta}{1-\beta} \sim 1 + \Gamma^2\delta\theta^2$ and $\Gamma\delta\theta \gg 1$. Therefore $L_\nu^{\mathrm{off}}/L_\nu^{\mathrm{on}} \sim f_{\mathrm{geo}}S^{-2}$, with $f_{\mathrm{geo}} = \Delta\phi/2\pi \sim 1/2$.

(iii) A significantly misaligned line of sight ($\delta\theta > \theta_0$): The whole emitting region has nearly the same Doppler factor and $\Gamma\delta\theta \gg 1$ still holds. Thus $d\Omega \sim \theta_0^2$ and one finds $L_\nu^{\rm off}/L_\nu^{\rm on} \sim (\Gamma\theta_0)^2 S^{-3}$. This case does not occur in the set-up of the present work.

Therefore, defining $S = \frac{1-\beta\cos\delta\theta}{1-\beta}$, the spectral luminosity transforms as the following for slightly misaligned lines of sight

$$\frac{L_\nu^{\rm off}}{L_\nu^{\rm on}} \sim f_{\rm geo} S^{-2}. \tag{A5}$$

With a similar reasoning, we have for bolometric luminosities

$$\frac{L^{\rm off}}{L^{\rm on}} \sim f_{\rm geo} S^{-3}. \tag{A6}$$

APPENDIX B: THE CONTRIBUTION OF LINE-OF-SIGHT MATERIAL TO THE PROMPT EMISSION OF SLIGHTLY MISALIGNED OBSERVERS

In Section 3, we adopt power-law structures in the jet material Lorentz factor and isotropic-equivalent prompt phase emitted energy as a function of latitude (equations 14 and 15), with parameters $\Gamma_j = 250$, $a = 8$, and $b = 6$. Our derivations suppose that the prompt emission – and therefore the ESD flux – detected by the misaligned observer comes from the material down the observer's line of sight. Here, we will show that this is the case for the adopted parameter values.

Adopting the notations $\epsilon_{\rm em, iso}(\theta)$ and $\Gamma(\theta)$ for the material's *emitted* isotropic-equivalent energy and Lorentz factor, the *observed* isotropic-equivalent energy at a viewing angle of θ_v is (e.g. equation 5 of Salafia et al. 2015)

$$E_{\rm obs,iso}(\theta_v) = \frac{1}{4\pi} \int \frac{\epsilon_{\rm em,iso}(\theta)}{\Gamma(\theta)} \delta(\theta, \phi, \theta_v)^3 d\Omega, \tag{B1}$$

where the integration runs over the jet structure, with the solid angle differential element $d\Omega = \sin\theta d\theta d\phi$. The δ is the relativistic Doppler factor:

$$\delta(\theta, \phi, \theta_v) = \frac{1}{\Gamma(\theta)(1 - \beta(\theta)\cos\chi)} \tag{B2}$$

and χ is the angular distance between the observer and the material at coordinates θ, ϕ:

$$\cos\chi = \cos\theta_v \cos\theta + \sin\theta_v \sin\theta \cos\phi. \tag{B3}$$

In Fig. B1 (left-hand panel; similar to Beniamini & Nakar 2019, Fig. 2 and Salafia et al. 2015, Fig. 4 and following), we plot $\epsilon_{\rm em, iso}(\theta)$ and $E_{\rm obs, iso}(\theta_v)$, normalized to the core emitted energy $\epsilon_{\rm em, iso}(\theta = 0)$, for the above-mentioned parameters and using equation (B1). In Fig. B1 (right-hand panel), we plot their relative difference, $2|E_{\rm obs, iso} - \epsilon_{\rm em, iso}|/(E_{\rm obs, iso} + \epsilon_{\rm em, iso})$. This function presents a singularity when the emitted and observed functions cross, as can be seen in the figure.

We find that for the slightly misaligned regime of viewing angles ($\theta_j \leq \theta_v \leq 2\theta_j$), we have $E_{\rm obs, iso}(\theta_v) = \epsilon_{\rm em, iso}(\theta_v)$, to within $\leq$ 18 per cent. In other words, the total observed energy at θ_v equals the emitted energy by the material down the line of sight, at $\theta = \theta_v$, and the prompt emission is dominated by the line-of-sight material. For larger $\theta_v \geq 2\theta_j$, we find that this approximation breaks down.
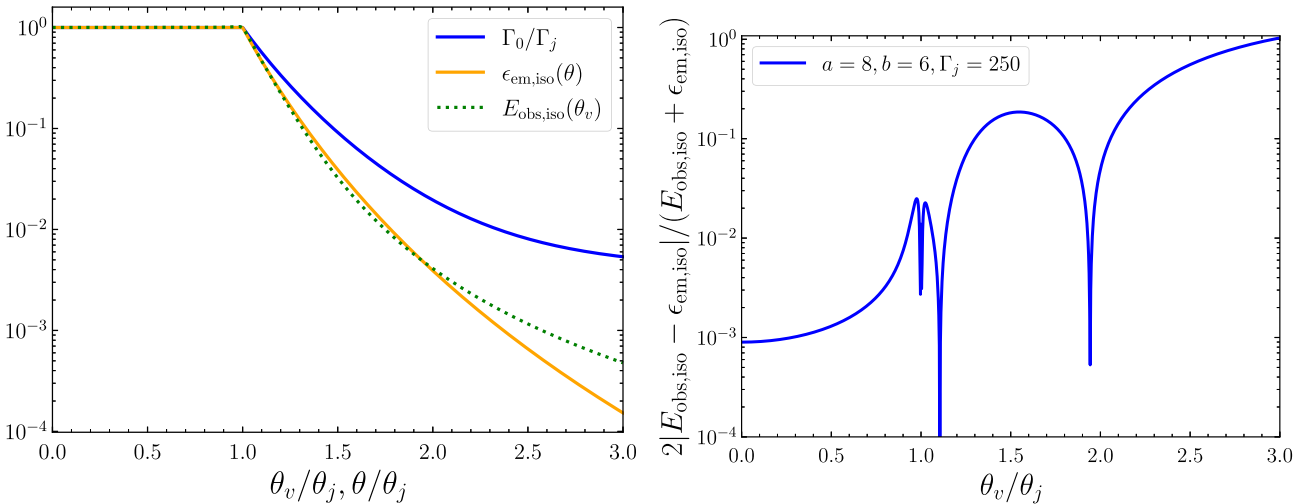

Figure B1. On the same axis for the latitude θ of the material in the structure and the viewing angle θ_v of the observer, we plot the prompt emitted energy $\epsilon_{\rm em, iso}(\theta)$ and the observed energy $E_{\rm obs, iso}(\theta_v)$ (left-hand panel), as well as their relative difference $2|E_{\rm obs} - \epsilon_{\rm em}|/(E_{\rm obs} + \epsilon_{\rm em})$ (right-hand panel). The relative difference function presents sharp dips whenever the emitted and observed energy functions cross. We can conclude from this figure that, for the adopted parameters ($\Gamma_j = 250$, $a = 8$, and $b = 6$, as in the text and the present appendix), the observed energy follows the emitted energy for viewing angles in the slightly misaligned regime, such that the observer's prompt emission can be said to originate from the material on their line of sight.

This paper has been typeset from a TEX/LATEX file prepared by the author.