ABSTRACT

Flaring activity following gamma-ray bursts (GRBs), observed in both long and short GRBs, signals a long-term activity of the central engine. However, its production mechanism has remained elusive. Here we develop a quantitative model of the idea proposed by Perna et al. of a disc whose outer regions fragment due to the onset of gravitational instability. The self-gravitating clumps migrate through the disc and begin to evolve viscously when tidal and shearing torques break them apart. Our model consists of two ingredients: theoretical bolometric flare lightcurves whose shape (width, skewness) is largely insensitive to the model parameters, and a spectral correction to match the bandpass of the available observations, that is calibrated using the observed spectra of the flares. This simple model reproduces, with excellent agreement, the empirical statistical properties of the flares as measured by their width-to-arrival time ratio and skewness (ratio between decay and rise time). We present model fits to the observed lightcurves of two well-monitored flares, of GRB 060418 and of GRB 060904B. To the best of our knowledge, this is the first quantitative model able to reproduce the flare lightcurves and explain their global statistical properties.

Key words: gamma rays: bursts — accretion, accretion discs — X-rays: general — black hole physics

1 INTRODUCTION

Over a decade of observations by the Swift satellite (Gehrels et al. 2005) have opened a new window on the phenomenology of gamma-ray bursts (GRBs). Arguably one of the most unexpected phenomena has been the discovery of erratic flaring activity in the X-ray luminosity of the sources. These X-ray flares have been observed over a timescale of seconds to even over a day after the prompt emission (e.g. Burrows et al. 2005; O’Brien et al. 2005; Margutti et al. 2010), hence overlapping with the timescale of the afterglow.

The standard model for the afterglow, that is the external shock model, predicts a lightcurve characterized by a powerlaw or broken powerlaws (e.g. Meszaros & Rees 1997), depending on the regime of observation. Several phenomena can cause departures from simple powerlaws, such as inhomogeneities in the external medium (Lazzati et al. 2002; Heyl & Perna 2003), refreshed shocks due to the collision of a late shell of plasma with the external shock material (Rees & Meszaros 1998), angular inhomogeneities in the fireball energy distribution (Nakar et al. 2003) and gravitational lensing (Loeb & Perna 1998). However, based on the observed properties of the flares, Lazzati & Perna (2007) demonstrated that—for at least the subset of them with a ratio between duration to arrival time \( \Delta t/t_p \lesssim 0.25 \)— inhomogeneities in the external shock cannot reproduce the observed properties of the flares. Rather, these must be the result of long-term activity of the GRB central engine.

Understanding the origin of the flares is of great interest because it would provide insight into the inner workings and physics of the GRB engines. The flares also provide an intriguing link between long and short GRBs (e.g. Margutti et al. 2011a), which are attributed to different progenitors, a collapsing star for the former (MacFadyen & Woosley 1999), and a compact binary merger for the latter (e.g. Narayan et al. 1992).

A number of ideas have been proposed over the years. For example, for the case of long GRBs, King et al (2005) suggested that the X-ray flares could be produced from the fragmentation of the collapsing stellar core in a modified hypernova scenario, while Lazzati et al. (2011) found that propagation instabilities of a jet piercing through a massive star might explain the properties of the subset of flares with...
2 fragmentation and clump formation in a disc

As discussed in Sec 1, the idea by PAZ that GRB flares may be due to viscously evolving blobs in a fragmented disc was partly motivated by numerous findings that the outer regions of a hyperaccreting disc are gravitationally unstable. The instability sets in when the Toomre parameter (Toomre 1964)

\[ Q_T = \frac{c_s \rho}{\pi G \Sigma} < 1, \]

where \( c_s \) is the sound speed, \( \rho \) the epicyclic frequency, \( G \) is the gravitational constant, and \( \Sigma \) is the disc surface density. Di Matteo et al. (2002) and Chen & Beloborodov (2007) found that, for accretion rates \( \dot{M} \sim 10 M_\odot\, s^{-1} \), the disc becomes unstable at radii \( R \geq 50 R_e \equiv R_{\text{inst}} \), where \( R_e = 2GM/c^2 \) is the Schwarzschild radius and \( c \) is the speed of light. The inner boundary of the unstable region, \( R_{\text{inst}} \), increases with decreasing accretion rate, and becomes \( \geq 1000R_e \) for \( \dot{M} \sim 0.1 M_\odot\, s^{-1} \). Recently, Liu et al. (2014) have extended the analysis of the gravitational instability to include the dependence of the Toomre parameter on the vertical structure of the disk. They confirmed that the instability sets in in the outer parts of the disk, and additionally they found that it is more easily produced further away from the midplane.

Once the disc becomes unstable, fragmentation into bound objects will occur if the local cooling time,

\[ t_{\text{cool}} < t_{\text{crit}} \approx 3 \Omega^2 \]

where \( \Omega_k \) is the local Keplerian angular velocity. When the above condition is satisfied, the maximum stress obtainable from gravitational instability produces insufficient heating to offset cooling (Gammie 2001; Rice et al. 2003). As discussed by PAZ, this condition is likely realized under the extreme conditions of GRB discs. A further in-depth analysis of cooling mechanisms in GRB discs was performed by Piro & Pfnal (2007). They demonstrated that, at disc radii for which \( Q_T \sim 1 \), photodisintegration is a very effective coolant that promotes fragmentation.

Once a disc becomes unstable and fragmentation ensues, the mass of a fragment is expected to be on the order of \( \Sigma \lambda^2 \), where \( \lambda \sim Q_T H \) is the wavelength of the fastest growing mode (Binney & Tremaine 1987), and \( H \) is the disk scale height. Hence, \( M_{\text{frag}} \sim \Sigma (Q_T H)^2 \). However, fragments are expected to merge and/or accrete until their tidal influence becomes strong enough to open a gap in the disc. This happens when the mass of the fragment has grown to

\[ M_{\text{frag}} \sim \left( \frac{H}{R} \right)^2 \alpha^{1/2} M_\star, \]

where \( M_\star \) is the mass of the central object (Takemichi et al. 1996), and \( \alpha \) the viscosity parameter in the disk (Shakura & Sunyaev 1973). Given that at the high accretion rates of GRB discs \( H \sim R \), and considering a range of masses between \( M_\star \sim 1.5 - 10 M_\odot \), it is reasonable to assume clumps with masses on the order of \( 0.05 - 1 M_\odot \).

In the following, under the assumption that the onset of the gravitational instability in the outer regions leads to self-gravitating clumps, we compute the time-evolution and the resulting luminosity of the accreting matter.
3 VISCOSOUS EVOLUTION OF SELF-GRAVITATING CLUMPS

A clump can be modeled as a sharp concentration of mass at some radius \( R \) (e.g. a delta function). It will arrive at the center (location of the compact object) and increase the luminosity at a timescale

\[
\tau_{\text{flare}} \sim \beta \nu \frac{R^2}{\nu},
\]

where \( \nu \) is the viscosity coefficient, and \( \beta \) is a dimensionless factor whose exact value depends on the initial profile and viscosity prescription (e.g. Milosavljevic & Phinney 2005, Tanaka & Menou 2010).

We compute the viscous evolution of the clump semi-analytically by solving the standard equation for axisymmetric accretion discs,

\[
\frac{\partial}{\partial t} \Sigma(R, t) = \frac{1}{R} \frac{\partial}{\partial R} \left( R^{1/2} \frac{\partial}{\partial R} \left( 3 \nu \Sigma R^{1/2} \right) \right).
\]

When the viscosity is a powerlaw of radius \( \nu \propto R^n \), the evolution of the surface density profile \( \Sigma(t; R) \) can be solved exactly by the integral

\[
\Sigma(t; R) = \int_{R_i}^{\infty} G(t; R, R') \Sigma(t = 0; R') \, dR',
\]

where the Green’s function \( G \) depends on the boundary conditions\(^2\). Solutions with inner boundary conditions (e.g. the no-torque condition at the the innermost stable circular orbit or a suppressed accretion rate due to gravitational torques or a magnetic field) imposed at the cylindrical coordinate origin \( R_i = 0 \) were derived by Lust (1952) and Lynden-Bell & Pringle (1974); these solutions have divergent disc luminosities. Solutions with \( R_i > 0 \), with \( t \to \infty \) disc luminosities matching those found in steady-state solutions, were derived by Tanaka (2011).

Once \( \Sigma(R; t) \) is known, the accretion power \( L_{\text{acc}}(t) \) due to the viscous spreading of the clump can be approximated as

\[
L_{\text{acc}}(t) \approx \int_{R_i}^{\infty} \frac{9}{4} \Sigma(R, t) \, \nu(R) \, \Omega_K(R) \, 2\pi R \, dR.
\]

Examples of viscosity prescriptions with the requisite form \( \nu \propto R^n \) include isothermal discs (\( \nu \propto R^{3/2} \), i.e. \( n = 3/2 \)) and advection-dominated discs (\( n \approx 1/2 \); \( c_s \sim v_K; H \sim R \); e.g. Tanaka 2013). Since the latter is more appropriate to the early accretion phases of interest here, we will adopt this scaling law.

The high-energy emission in GRBs is believed to be produced within a relativistic outflow (at the distances at which it becomes optically thin), and hence a crucial ingredient of GRB models is the ability to launch such jets. Two main mechanisms have been proposed: neutrinos (Popham,

\[^2\] The main advantage of the Green’s function method is that it allows for the calculation of disk evolution and accretion power with a single integration, for arbitrary initial surface density profiles.
Woosley & Fryer (1999, Ruffert & Janka 1999) and magnetic fields (Blandford & Znajek 1977; Blandford & Payne 1982). As it is found that neutrinos might play an important role only at the highest mass accretion rates (e.g. Janiuk et al. 2007, Janiuk et al. 2013, Just et al. 2016), we will not consider this mechanism here. Instead, we will focus on scenarios where the rotational energy of a spinning BH is extracted by a large scale magnetic field threading the disk (Blandford & Znajek 1977). This mechanism is commonly considered as the most viable for launching jets, and predicts a jet power that is directly proportional to the mass accretion rate$^2$ (Blandford & Znajek 1977, Krolik & Piran 2011, 2012; see also Kumar, Narayan & Johnson 2008). The jet power is then converted into radiation farther along its path, e.g. through collisions of shells with different velocities (Burrows et al. 2005; Fan & Wei 2005; Zhang et al. 2006, Kumar & Zhang 2015). As long as the radiative mechanism maintains a constant efficiency throughout the duration of an individual flare, we can assume that the bolometric luminosity of the flare will be $L_{\text{bol}} = f_{\text{rad}} L_{\text{acc}}$, where $f_{\text{rad}}$ - the efficiency of energy conversion from mass accretion into radiation - is a constant for that event, but may change among different events.

4 Flares Due to Self-Gravitating Clumps: Evolution Under Different Physical Conditions

Once a clump has been formed, it will eventually evolve viscously according to Eq. (6). In this work, we envisage two main scenarios for the time when viscous spreading begins.

4.1 Early Flares: Prompt Viscous Spreading

The first scenario we consider is one in which clumps form during the time that the prompt emission is produced, in regions of the disc where viscous stresses are strong enough to shred the clumps right away. In this case, viscous evolution begins without significant delay with respect to the GRB prompt. The condition for this to happen will be discussed in more detail in Sec. 4.2.

Examples of solutions to Eq. (6) are shown in Fig. 4 for a range of values of the clump mass, the viscosity parameter, the disc radius at which the clump evolution starts ($R_{\text{in}}$), and the mass of the accreting object. To include the possibility that the central engine is a neutron star, we also considered a mass value of 1.5 $M_{\odot}$. The figure shows the bolometric luminosity, computed via using Eqs. (6) and (7).

We note that, as expected, the mass of the clump essentially determines the magnitude of the flare, while the initial radii, the viscosity parameter, and the mass of the central object influence the peak time and the width. These then affect the peak luminosity, since the total radiated energy $E \propto M_{\text{cl}}$ is the same for a given clump mass. More specifically, we find that the peak luminosity of the flares scales with their initial position as (roughly) $R_{\text{in}}^{-1.6}$, and the luminosity scales with the peak time as (roughly) $t_{\text{p}}^{-0.96}$.

These bolometric lightcurves are characterized by a ratio of the flare “width” to peak time on the order of $\Delta t / t_{\text{p}} \approx 4.3$. Here the width $\Delta t$ is defined as the interval between the times $t_1$ and $t_2$ (the former during the rise and the latter during the decay) at which the flux is 1/e of the peak flux. If, in addition, we define the rise time as the difference $t_2 - t_1$ and the decay time as $t_d - t_2$, then the bolometric lightcurves are also characterized by a ratio $t_d/t_r$, which measures the asymmetry (or skewness) of the curve. We systematically find $t_d/t_r \approx 5.7$. These quantities are defined in the same way as the observational characterizations of Margutti et al. (2010), to allow for direct comparisons (see Sec. 5).

We note and emphasize an important property of our model lightcurves: the ratios $\Delta t/t_{\text{p}}$ and $t_d/t_r$ are almost insensitive to the value of model parameters. The robustness of the light curve shape at early times is due to the fact that as long as the initial position of the clump is far away from the inner boundary condition imposed at $R_{\text{ISCO}}$, the early viscous evolution simply scales with the local viscous time profile ($\propto R^2/\nu(R) \propto R^{2-n}$) of the clump. At late times, $L_{\text{bol}}(t) \propto t^{-4/3}$ if $n = 1/2$ (i.e. if $H/R \sim$ constant), with the decay powerlaw depending on $n$ (see Tanaka 2011).

Increasing the initial distance of the clump increases the timescales of the early-time evolution by a factor 2$^n$ but otherwise the surface density evolution is the same. Similarly, $L_{\text{bol}}(t)$ simply scales with the normalization of $\Sigma$, so changing the clump mass in our model does not change the shape of $L(t)$.

We have also verified that replacing the $\delta$-function initial profile with Gaussian or step-function annuli does not strongly affect the shape of the curve. Hence, the shape of the bolometric light curves can be considered as distinctive, predictive signature of the model.

These properties will be discussed in more detail and compared to the data in Sec. 4.2.

4.2 Late Flares: Delayed Viscous Spreading

This is the case in which the viscous evolution of clumps begins with a significant delay with respect to the prompt emission. In general, there are two reasons for such a delay to occur. First, clumps may form soon after the prompt, as in sec. 4.1, but in outer regions of the disc where they can survive viscous stresses for a while before eventually migrating toward the inner region of the disc where they are shredded. Second, disc fragmentation may last for a while, hence clumps may be formed with some delay with respect to the initial prompt emission.

As time progresses after the prompt phase, the accretion rate drops, and hence the radius $R_{\text{out}}$ below which the Toomre parameter $Q_P < 1$ increases, $R_{\text{out}} \propto M^{-2/3}$. The accretion rate during the flare phase can be estimated assuming that it tracks the mean observed luminosity (Lazzati et al. 2008; Margutti et al. 2011b). From a sample of 44

Note that this is not true generally if the viscous properties of the disc scale non-linearly with $\Sigma$. In that case, changing the clump mass or the initial clump location could lead to different light curve shapes.

3 Apart, possibly, from the very early phases associated to the prompt emission (Tchekhovskoy, Narayan & McKinney 2010).
bursts, Margutti et al. (2011b) estimated a decline with an average power of $t^{-2.7}$. Hence, during this phase, the radius at which $Q_T = 1$ expands outward as $R_{Q_T=1} \propto t^{1.8}$. This means that, the later the times, the larger the disc radii at which clumps can form. Clumps that form very far out in the disc, in regions that are very gravitationally unstable and in which viscous torques are rather weak, are likely to remain bound and self-gravitating for some time, migrate inward, and only start spreading when shear and/or tidal forces become comparable to their self-gravity.

The shear force per unit length acting on a fluid element at radius $r$ is $F_v = \nu \Sigma r \frac{d\Omega}{dr}$. A clump of linear size $\ell$ and mass $M_{cl}$ will thus be subject to the shear force per unit mass

$$F_v = \ell F_v = \nu \Sigma r \frac{d\Omega}{dr} = \frac{\ell M_{cl}}{2\pi M_{cl}} \Omega, \quad (8)$$

where we made use of the relation $M = 3\pi \nu \Sigma$, valid in a steady-state disc, and of the fact that $\Omega \propto r^{-3/2}$ for keplerian rotation.

Given the self-gravity force per unit mass of the clump, $F_{SG} = \frac{GM_{cl}}{r^2}$, and the tidal force due to the central object, $F_T = \frac{GM}{r^2} \left(\frac{\ell}{r}\right)$, viscous spreading of the clump will start roughly when the ratio $V = \frac{F_v + F_T}{F_{SG}} > 1$. This condition can be re-cast as

$$\frac{M}{M_{cl}} \left(\frac{\ell}{R_s}\right)^3 x^{-3} \left[1 + \frac{\sqrt{2} R_s}{c} \frac{M}{M_{cl}} x^{3/2}\right] > 1 \quad (9)$$

where we have defined $x \equiv \ell / R_s$. For a given central object and clump mass and size, this condition defines a boundary in the $r$ vs. $M$ plane between (inner) regions where clumps form and start spreading right away, and (outer) regions where newly formed clumps remain self-bound. In the latter case, they will have to migrate inwards before being viscously/tidally shredded. For an order of magnitude estimate, we assume $M_s = 10 M_{\odot}$ and rescale the clump size based on the local Jeans length, $\lambda_J = c_s^2 \frac{\pi}{G \rho} \sim 2 \times 10^8 \left(\frac{M}{100 M_{\odot}}\right)^{-1/2} \text{cm}$, where the disc sound speed $c_s \approx H \Omega_K \sim 0.5 c_k$ is typical for a slim disc and we have approximated the density profile for $x > 100$ as $\rho \sim 10^9 \left(\frac{M}{100 M_{\odot}}\right) \left(\frac{x}{100}\right)^{-2} \text{g cm}^{-3}$ (Di Matteo, Perna & Narayan 2002). Therefore, we assume a typical clump size $\ell \sim 10^8 \text{cm}$. Eq. 9 leads us to conclude that the disc radius at which viscous/tidal disruption occurs is in the range $200 - 1000 R_*$, depending on $M$ and on the specific values of the relevant physical parameters.

A thorough discussion of all possible behaviors resulting from Eq. 9 is beyond the scope of this paper. Such an analysis would require detailed numerical simulations of the process of clump formation, cooling and fragmentation in a rapidly evolving accretion disc, where $M$ can drop by 2-3 orders of magnitude in hundreds of seconds.

For the discussion presented in Sec. 5, what matters is that clumps may form at some outer radius $R_0$, and then migrate inwards till a radius $R_{in}$ before they begin to evolve viscously. The migration timescale can be comparable to the viscous time, if the local gravity is dominated by the disc. On the other hand, if the clumps are massive enough and dominate the local potential, they will migrate inward on a timescale $t_{mig}$ which is longer than the viscous timescale by a factor on the order of $(M_{cl}/M_{disc}) > 1$, where $M_{disc}$ is the exterior disc mass, and details depend on the precise radial dependence of the density profile of the unperturbed disc (Syer & Clarke 1995). At later times, when fallback from the envelope of the star has terminated and hence the remaining disc is only depleting, the condition $M_{cl}/M_{disc} > 1$ may be verified, further lengthening the travel time of the clump.

In summary, as the mean accretion rate in the disc drops, clumps form farther out and may travel in the disc before beginning to viscously spread. Under these conditions, their travel time may be comparable to, or longer than the viscous timescale at $R_0$. Additionally, as the disc expands and its conditions change with time, new clumps in the outer regions of the disc can form some time after the prompt emission. In the following, let us call $t_{off}$ the offset time between $t = 0$ (defined as the time of the GRB trigger) and the time at which viscous spreading begins. Again, this offset could be due to a long migration time, or to late-born clumps, or to a combination of these two factors.

Figure 2 compares the observed lightcurves for a clump of identical properties (i.e. mass, radius where spreading begins) but only differing for the time $t_{off}$ at which spreading begins. We note that the shape of these curves is independent of $t_{off}$ — the decay time and the rise time do not vary, hence the ratio $t_{off}/t_r$ stays constant. The duration $\Delta t = t_{off} + t_r$ is also unchanged; however, since the flares arrive at later times, the ratio $\Delta t/t_p$ gets smaller for larger $t_{off}$. These properties will be further discussed and compared to the observations in Sec. 5.
5 COMPARING OUR MODEL AGAINST OBSERVATIONS

5.1 lightcurve shape and $\Delta t/t_p$ correlation

As already mentioned in Sec.4, an important feature of our model lies in the fact that the bolometric lightcurve produced by the viscous spreading of the clump has a well-defined shape, which is only marginally affected by changing the values of the physical parameters.

We have studied how the values of $t_p$, $t_1$ and $t_2$ change as $\alpha$, $R_{in}$, $M_*$ and $R_{ISCO}$ are varied within physically plausible ranges. Our main conclusion is that, while each individual timescale changes according to the numerical value of each of the physical parameters (cfr. Fig. [1]), their ratios remain remarkably constant. In particular, the “asymmetry parameter” $t_d/t_s \approx 5.7$ and the “width parameter” $\Delta t/t_p \approx 4.3$ are insensitive to the value of $R_{in}$, $M_*$ and $R_{ISCO}$. Only increasing the powerlaw index $n$ of the viscosity law from 0.5 (an advection-dominated disc with $H/R \sim 1$) to 1.5 (an isothermal disc) can appreciably affect the lightcurve shape, producing narrower peaks with somewhat smaller asymmetry. However, as discussed above, GRB discs are expected to be in the former regime.

In order to test our model predictions against actual flare data, we consider the sample of flares with $t_p \lesssim 10^3$ s from Margutti et al. (2010), together with the sample of long-lag, wide prompt pulses from Norris et al. (2005). We include this additional sample because X-ray flares can be effectively fit with the same functional shape of the long-lag, wide prompt pulses (Margutti et al. 2010), allowing a direct comparison of their temporal properties. Note that in most bursts, the prompt emission consists of a complex forest of overlapping pulses with different temporal and spectral properties (e.g. Norris et al. 1996). Interestingly, Norris et al. (2005) demonstrated that prompt pulses in GRB with long temporal lag between the high-energy and low-energy photons are spectrally and temporally distinguished from those in bright GRBs: the pulses are wider, with lower peak fluxes, harder low-energy spectra and softer high-energy spectra. We refer to these as ‘prompt flares’ hereafter.

Fig. [2] shows the flare width versus peak time (left panel) and the decay time versus rise time (right panel) for the full sample described above. It is interesting to note that the ‘prompt flares’ appear as a somewhat distinct population than that of the ‘standard’ flares when it comes to $\Delta t/t_p$: they appear to have a similar slope for the $\Delta t/t_p$ correlation, but the latter flares have an offset, signaling a later arrival time but otherwise similar underlying physical properties. The similarity between the underlying phenomena is strengthened by the fact that no separation appears in the two populations when looking at the $t_d$-$t_s$ plane: the shape of the flares, as measured by their asymmetry parameter, appears to be independent of their arrival time (with some scatter, which will be discussed in the following).

The theoretical predictions of our model can now be directly compared with the data of Fig. [2]. As already noted, the $t_d$ vs. $t_s$ relation is simply given by a straight line with slope $\approx 5.7$. Remarkably, this line appears to match the upper edge of the scattered distribution of flares in Fig. [2] (right panel, solid line). Additionally, our model can be represented by another line with slope 4.3 in the $\Delta t$ vs. $t_p$ plane, which appears to match very well the distribution of “early flares”.

Again, the points show some scatter, but it is remarkable that our model, with its very rigid predictions, falls exactly in the middle of the strip of data points.

As noted above, while all data points in the $t_d/t_s$ plane appear to lie along the same distribution, with a scatter corresponding to a ratio $t_d/t_s \sim (1.5 - 6)$, the $\Delta t - t_p$ data show a large population of “late flares” that appear to be offset with respect to our model expectations. These can be fit in the proposed picture, however, if there is an appreciable delay (i.e. what we call $t_{off}$, cfr. Sec.4.2) between the “time zero” of the observations and the time at which they started spreading viscously in the disc. As argued in Sec.4.2, this can indeed be the case if the clumps form at some outer radius $R_0$, and migrate to some radius $R_{in} < R_0$ before they can begin to shred due to tidal and/or viscous torques. Or alternatively, they simply form at some later time after the prompt emission.

Within this scenario, the early flares ($t_p \lesssim 10 - 30$ s) would thus correspond to clumps formed in the inner region of the disc, when $M$ is the highest and viscous spreading of the clumps begins very close to their formation site; later clumps will instead be formed in the outer disc, due to a strong decrease of $M$, where their self-gravity initially resists viscous stresses. Such self-bound clumps would migrate inwards without spreading, until reaching smaller radii where viscous evolution eventually takes over: at this point, they would display the same evolution as the early ones, just with a time delay $t_{off}$. In particular, their shape parameter $t_d/t_s$ would remain the same, while the ratio between the width and arrival time, $\Delta t/t_p$, would be decreased by a factor $\sim t_{off}/t_s$, giving rise to a line parallel to that of early flares in the $\Delta t$ vs. $t_p$ plane.

This interpretation is consistent with the fact that the first $\sim 10 - 20$ s of the emission roughly correspond to the typical duration of a long GRB, during which the collapsing envelope of the star maintains the accretion rate at a high and roughly (on average) constant value. Once the fallback has ended, the accretion rate drops abruptly and decays as a powerlaw (MacFadyen & Woosley 1999; Lazzati et al. 2008; Perna et al. 2014). With a lower accretion rate, clumps form at increasingly larger radii, and, as argued above, they need to migrate inwards before they can be shredded by the viscous forces within the disc.

Last, it is useful to derive, within the framework of our model, a relation for the width $\Delta t$ versus $t_p$, inclusive of the possibility of a delayed viscous spreading $t_{off}$. Generally, we can write $t_p \approx t_v + t_{off}$. Our bolometric light curves predict $\Delta t/t_p \approx 4.3$ in case of prompt spreading where $t_p \approx t_v$. Therefore, if there is a delayed spreading, we can write $\Delta t \approx 4.3 t_p - 4.3 t_{off}$. In the $\Delta t - t_p$ plane these are parallel lines of slope 4.3, and offset determined by $t_{off}$. The group of late flares is the one for which $t_{off} \gg t_v$; hence, since $t_{off}$ is likely to vary from burst to burst, our model predicts a larger scatter for the “late” flares than for the “early” ones, for which $t_p \approx t_v$. If, for some bursts, $t_{off} \sim t_v$, then these would represent a subset of flares with intermediate values of $\Delta t/t_p$; however, on simple statistical grounds these would be expected to be less numerous. The data in the left panel of Fig. [3] supports both these predictions: the late flares display


Figure 3. A comparison of the flare data from the Margutti et al. (2010) sample with the predictions of our model. Left: Width-to-peak time correlation. Right: Decay time versus rise time for each flare. The solid lines in both panels indicate the (highly constrained) model predictions from the bolometric lightcurves. Spectral corrections due to the finite band width introduce scatter and decrease the ratios $\Delta t/t_p$ and $t_d/t_r$ (as indicated by the arrows), albeit the latter is much more sensitive to spectral corrections than the former. As an example, the dashed lines indicate the spectral correction obtained with the measured parameters of a specific flare, that of GRB 060904B (see Sec.5.2).

Figure 4. Left panel: The bolometric lightcurve $L_{bol}(t)$ (blue line), together with the 0.3-10 keV lightcurve $L_X(t) = \epsilon_X(t)L_{bol}(t)$ (red line) for the bright flare in GRB 060904B. The inset shows the time-dependent X-ray efficiency $\epsilon_X(t)$ derived from the observed spectrum of this flare. Right panel: The 0.3-10 keV luminosity of the flare, superimposed to the data. Note that this is not a formal fit since, once the peak time of the lightcurve is matched with a choice of model parameters, the shape of the bolometric lightcurve is highly constrained, and the spectral correction is unique to this flare. For these plots, we adopted $M_* = 10M_\odot$, a viscosity parameter $\alpha = 0.1$ and power-law index $n = 0.5$ for the viscosity radial profile (see Sec. 3).

significant more scatter than the early ones, and there is a paucity of data in the intermediate region.

5.2 Spectral Effects: Modeling of the Flares in GRB 060904B and GRB 060418

The next step in comparing our model predictions to observations calls for the conversion from bolometric to X-ray luminosities (in the 0.3-10 keV range, where most flares have been observed), with the goal of testing whether the model is able to: i) Explain the scatter, in addition to the slope, in the statistical comparison with the data of Fig. 3; ii) reproduce the observed lightcurves of individual flares.

To this end, we will rely on the observed spectra of two particularly bright flares, GRB 060904B and GRB 060418, which were studied in detail by Margutti et al. (2010). These authors showed that flare spectra can be well fit by a Band function (Band et al. 1993), the spectral parameters of which change with time. In both flares the peak energy, $E_p$, has an exponential decay after the flux maximum, with the same timescale of the flux decay. In addition, Table 3 of Margutti et al. (2010) shows that the two spectral indices $\alpha$ and $\beta$ are also evolving on the same timescale, and settle at later times at some constant values. In light of these results, for both flares we have adopted the following analytical formulae to

Note also that, due to the several tens of seconds for repointing a satellite in X-rays ($\gtrsim 60$ s for Swift) after the $\gamma$-ray emission has faded, there is an observational bias against observing flares with several tens of seconds duration.
approximate the evolution of their spectral parameters

\[ E_p(t) = \begin{cases} 
E_{p,i} & \text{if } t \leq t_{\text{max}}, \\
E_{p,i} e^{-(t-t_{\text{max}})/\tau} & \text{if } t > t_{\text{max}}, 
\end{cases} \tag{10} \]

and

\[ \begin{align*}
\alpha &= \alpha_{\text{late}} + a_1 e^{(t-t_{\text{max}})} \\
b &= \beta_{\text{late}} + b_1 e^{(t-t_{\text{max}})}
\end{align*} \tag{11} \]

where the numerical coefficients in the two cases are

| flare     | \(E_p[\text{keV}]\) | \(\tau[\text{s}]\) | \(\alpha_{\text{late}}\) | \(\beta_{\text{late}}\) | \(a_1\) | \(b_1\) |
|-----------|---------------------|-------------------|-----------------|-----------------|--------|--------|
| GRB 060904B | 7.2 | 20 | -1 | -3 | +0.5 | +0.5 |
| GRB 060418 | 5.7 | 20 | -1 | -2.5 | +0.1 | 0 |

By means of these formulae we can calculate, as a function of time, the fraction of the total energy that was emitted in the 0.3-10 keV, \(\epsilon_X(t) = \left(\int_{0.3}^{10} E \, N(E) \, dE \right) / \left(\int_{0.01}^{150} E \, N(E) \, dE\right)\).

With this coefficient at hand, we can transform our theoretical bolometric lightcurves into X-ray lightcurves, \(L_X(t) = \epsilon_X(t)f_{\text{rad}}L_{\text{acc}}(t)\). The total efficiency of conversion of accreted mass into radiation is estimated to be typically of the order of a few percent (Giacomazzo et al. 2013 and references therein): as a reference value, we have adopted \(f_{\text{rad}} \sim 0.01\).

We stress how, by using an observationally-determined spectral efficiency, our approach does not rely on any specific assumption about the (rather uncertain) emission mechanism. However, we note that an efficiency that drops with time, as inferred by our modelling (see Figs. 4 and 5), is consistent with the “curvature effect” if the emitting region has an accelerated bulk relativistic motion (Uhm & Zhang 2015, 2016).

For completeness, some clarification about the value of \(t_{\text{max}}\), the time of the lightcurve maximum, is in order here. Both events considered here belong to the group of “late flares”, with \(t_f > 100 \text{ s}\), most likely reflecting the enhanced migration time prior to viscous spreading as discussed in Sec. 5.1. Hence, we calculated the spectral efficiency, \(\epsilon_X(t)\), for bolometric lightcurves that have a rise time matching the observed \(t_f\) of the two flares under study (since \(t_f\), unlike \(t_p\), is unaffected by the absolute value of the arrival time as measured from the trigger). The value of \(t_{\text{max}}\) was then set to correspond to the maximum of the bolometric lightcurves and, from these, the (0.3-10 keV lightcurve were calculated in the way described above. Finally, we introduced in all the functions a time offset as an arbitrary parameter, that simply shifted the calculated X-ray lightcurves until they matched the peak of the observed ones (while the shape was fixed).

To convert the (0.3 - 10) keV fluxes to the corresponding luminosities, we used the measured redshifts for the two flares (Margutti et al. 2010) and the “concordance” values for the cosmological parameters, \(\Omega_M = 0.27, \Omega_\Lambda = 0.73\) and \(H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}\). As for the (unknown) beaming factor of the received radiation, \(f_b\), we adopted a typical value \(\sim 0.01\) (e.g., Frail 2001).

Our results are shown in Figs. 5 and 4 for each flare, the left panel shows the bolometric lightcurve (blue), calculated for chosen values of the physical parameters, and the corresponding X-ray lightcurve (red) derived by adopting the observed spectral parameters (see above). In the inset, the corresponding spectral efficiency as a function of time is also shown. The right panel of each figure shows the superposition of the calculated X-ray lightcurve (red) to the data.

We stress that the curves do not represent formal fits: they were obtained by choosing specific values of the physical model parameters and adopting the observed spectra of the two flares. Nonetheless, we emphasize an important result: since the shape of the bolometric lightcurve is highly constrained (as discussed in Sec.4 and 5.1) and depends very little on the model parameters, the actual, observable shape of the light curve becomes very sensitive to the actual spectrum of the flare. Here we have selected two flares: one with a high spectral correction (GRB 060409B), and another with a marginal spectral correction (GRB 060418). The data are reproduced remarkably well by the model simply by modeling each flare with its own observationally-derived spectral efficiency. We found that adopting an “average spectral correction” failed to produce satisfactory fits of individual lightcurves. We have found this result to be an especially attractive feature of our model.

Note that, for fixed \(M_i\) and \(\alpha\), the value of the initial radius \(R_0\) is essentially determined by the rise time of the flare, leaving the mass of the clump as the only parameter.
sensitive to $f_b$ and $f_{\text{rad}}$. Our results for $M_{\text{cl}}$ were obtained for $f_b = 1$, and scale roughly linearly with $f_{\text{rad}}$.

Finally, our exploration of the spectral effects on the bolometric light curves has shown that, as a general feature, spectral corrections via a time-dependent Band function tend to make the lightcurves narrower. However, the correction to $\Delta t/t_p$ is smaller than that to $t_d/t_e$, since the time-dependent efficiency $c_X(t)$ drops with time and hence it influences especially $t_d$ (see insets in the left panels of Figs. 5 and 6). From a statistical point of view, this means that the theoretical, bolometric predictions for $\Delta t/t_p$ and $t_d/t_e$ (solid red lines in Fig. 4) provide an upper value for those quantities (modulo a small theoretical scatter with changes in the model parameters). On the other hand the spectral corrections, which differ from flare to flare, introduce a scatter in these quantities, which is especially pronounced for $t_d/t_e$, and goes in the direction of smaller values for the reasons explained above. As an example, the dashed lines in the two panels of Fig. 5 show the values of $\Delta t/t_p$ and $t_d/t_e$ using the spectral efficiency derived for GRB 060418B.

In summary, the statistical properties of the flares are naturally reproduced by our model.

6 SUMMARY AND CONCLUSIONS

We have built a quantitative model for the idea suggested by Perna et al. (2006) to explain the flares seen in both the long and the short GRBs. An hyperaccreting disc becomes gravitationally unstable and fragments in its outer regions. With a sufficiently efficient cooling, the fragments become self-gravitating. The matter clumps start migrating inwards and, when they reach a location at which viscous and/or tidal shears are able to tear them apart, they begin spreading and accreting viscously.

By computing the time-evolution of the matter in the clumps and the associated accretion luminosity, we have identified several robust features of this model:

- If the clumps begin to spread close to the location where they were formed, the ratio between their duration and arrival time is a robust prediction of the model, $\Delta t/t_p \approx 4.3$, with an insignificant sensitivity to all the model parameters within reasonable ranges. If the clumps start spreading viscously after some migration time within the disc, then the ratio $\Delta t/t_p$ decreases; however, the absolute value of the width remains roughly constant.
- The asymmetry of the curve, as measured by the ratio between the decay and the rise time, is remarkably independent of the model parameters, and has a roughly constant value of $t_d/t_e \approx 5.7$; it is the same whether the clump arrives after a migration time delay or not.
- Spectral corrections, due to the fact that observations are performed in a finite energy band, play an important role in modifying the shape of the time-dependent bolometric luminosity predicted by our model. In particular, the fast decline of the flare lightcurves does not directly track the decline of the accretion power, but is best explained in terms of a marked spectral evolution. This conclusion is consistent with recent findings on the effect of accelerated relativistic bulk motion in the emitting region (Uhm & Zhang 2016).
- By using the observationally determined spectra of several flares, we have computed the corresponding time-dependent spectral corrections to the measured luminosity, and demonstrated that the spectral corrections only have the effect of decreasing the width $\Delta t/t_p$ and the ratio $t_d/t_e$ compared to the same quantities calculated without the spectral correction. However, the correction on $t_d/t_e$ is larger than for $\Delta t/t_p$.
- A comparison of the model predictions with the data shows a strong agreement which is even more remarkable given the fact that model is very constrained and quite insensitive to the model parameters. Our predictions naturally match the data, they are not fit to the data. More specifically, the predictions for $\Delta t/t_p$ pass through the data, and there is a relatively narrow scatter given by spectral corrections. On the other hand, the theoretical line for $t_d/t_e$ provides an upper limit to the data, and the relatively large scatter at lower values is due to the more pronounced sensitivity to spectral corrections.
- While we predict that $t_d/t_e$ should be the same independently of whether the clump started to spread right away or migrated first, on the other hand the ratio $\Delta t/t_p$ decreases if there has been a migration time $t_{\text{off}}$ before viscous spreading begins. The data suggests two groups, one with $t_{\text{off}} \approx 0$, and the other with $t_{\text{off}} \sim 2$. We identify the former group with clumps formed during the prompt phase, when the conditions in the disc are roughly steady-state, while the latter with the post-prompt phase, when the accretion rate begins to rapidly drop, and the radius beyond which the disc can be unstable migrates outward. Clumps formed in the outer, lower density regions of the disc need to migrate inwards before they can begin to be shred.

In addition to explaining the main properties of the flares from a statistical base, we also modeled two specific flares with good data coverage. For each of them we corrected the bolometric lightcurve (which, again, has a well-constrained shape) with the corresponding time-dependent efficiency, determined for each of them from their own observationally determined time-dependent spectra. Once each is corrected for their own spectral efficiency, the theoretical lightcurve matches the data remarkably well.

ACKNOWLEDGEMENTS

We thank Andrew MacFadyen and Riccardo Ciolfi for valuable discussions, and Bing Zhang and the anonymous referee for useful comments on the manuscript. RP acknowledges support from NASA-Swift under grant NNX15AR48G, and from the NSF under grant AST-1616157.

REFERENCES

Band, D., Matteson, J., Ford, L., et al. 1993, ApJ, 413, 281
