

## PROMPT GeV EMISSION FROM RESIDUAL COLLISIONS IN GAMMA-RAY BURST OUTFLOWS: EVIDENCE FROM *FERMI* OBSERVATIONS OF GRB 080916c

ZHUO LI<sup>1,2</sup>

<sup>1</sup> Department of Astronomy, Peking University, Beijing 100871, China

<sup>2</sup> Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China  
Received 2009 February 20; accepted 2009 December 8; published 2010 January 4

### ABSTRACT

The  $\gamma$ -rays from gamma-ray bursts (GRBs) are believed to be produced by internal shocks driven by small timescale,  $\sim 1$  ms, variation in the GRB outflows, and a pair-production spectral cutoff is generally expected around the GeV range. However, the observed optical flashes accompanying GRBs suggest that the delayed residual collisions due to large timescale variation continue to accelerate electrons. We show here that the inverse-Compton (IC) scattering of the prompt  $\gamma$ -rays by these residual internal shock electrons leads to a high-energy emission beyond the previously thought spectral cutoff, in agreement with the previous detections of GeV photons by EGRET in several GRBs in conjunction with MeV emission. We expect a spectral break due to the transition from the primary to residual internal shock emission at the previously thought spectral cutoff and expect systematic time delays of high-energy photons relative to MeV emission, the discovery of which would provide stringent constraint on the outflow properties, but requires large enough collection of high-energy photons by, e.g., *Fermi* and *AGILE* satellites. The recent *Fermi*-detected bright GRB 080916c unambiguously shows the shifting of the prompt emission toward later times as the photon energy increases. The second-scale shifting at  $> 100$  MeV is much longer than the MeV variability time, as predicted in the residual collision model. The observations imply that there should be emission above 70 GeV in the source frame, which may not be produced by primary internal shocks but by IC emission in residual collisions. With the method involving time delays of high-energy emission, the bulk Lorentz factor of GRB 080916c is determined to be  $\Gamma \sim 300$ .

**Key words:** acceleration of particles – gamma-ray burst: general – magnetic fields – shock waves

### 1. INTRODUCTION

The prompt MeV  $\gamma$ -rays from a gamma-ray burst (GRB) are well believed to be produced by an unsteady outflow which causes internal collisions between different parts with various velocities, leading to kinetic energy dissipation (Paczynski & Xu 1994; Rees & Meszaros 1994; see, however, Lyutikov & Blandford 2003; Narayan & Kumar 2009). The internal shock model can naturally explain both the non-thermal spectra and the complicated light curves of GRBs. The observed temporal variabilities of  $\gamma$ -ray emission are believed to be reflecting the activities of the central engine (Sari & Piran 1997; Kobayashi et al. 1997). The internal shocks are expected to generate/amplify magnetic field and accelerate electrons, leading to MeV  $\gamma$ -rays by synchrotron emission (see, e.g., Waxman 2003 for a review).

GRBs are mainly observed in MeV range, the properties of high-energy, say,  $> 100$  MeV, emission are not well understood, which, however, might be very helpful in constraining the physics of the GRB emission region. For example, the observed  $> 100$  MeV photons in conjunction with MeV emission by EGRET in several GRBs, suggesting that they can avoid the  $\gamma\gamma$  absorption, have led to the conclusion that the GRB outflow must be relativistically expanding with a Lorentz factor of  $\Gamma \gtrsim 100$  (e.g., Baring & Harding 1997; Waxman 2003). As the development in high-energy  $\gamma$ -ray observations, e.g., *AGILE* and *Fermi* are being well operated, there are great interests on detecting the high-energy pair-production spectral cutoff in order to constrain the size and Lorentz factor of GRB emission region (e.g., Baring 2000; Lithwick & Sari 2001; and recent detailed consideration by Gupta & Zhang 2007, 2008; Granot et al. 2008; and Murase & Ioka 2008).

However, a simple spectral cutoff may not exist. Li & Waxman (2008, hereafter LW08) had noticed that the frequently observed prompt optical emission (Akerlof et al. 1999; Blake et al. 2005; Vestrand et al. 2005, 2006; Yost et al. 2007) is above the expected synchrotron-self-absorption limit and suggests a relatively large size of optical emission region, compared to that of MeV emission. Actually in the context of internal shock models one would expect delayed collisions in the outflow following the small radius collisions driven by smallest timescale variation of the outflow properties, and these delayed collisions can naturally account for the relatively bright optical emission (LW08). The recently detected optical flash from the naked-eye GRB 080319b (Bloom et al. 2009; D’Elia et al. 2009; Racusin et al. 2008; Wozniak et al. 2009) appears to vary rapidly in times and its temporal profile is correlated to the MeV emission in second scales (Beskin et al. 2009), supporting that the prompt optical emission in GRBs is produced by internal shocks within the outflow, i.e., synchrotron emission from residual collisions (Z. Li et al. 2010, in preparation). The electrons in residual collisions mainly cool by inverse-Compton (IC) scattering off the MeV photons, which produces high-energy emission at large radii where the optical depth due to pair production is reduced (LW08). We consider here this high-energy emission and show that it may “smear” the previously thought pair-production spectral cutoff. But a spectral turnover is still expected, which may be observed by *Fermi* and *AGILE* although more difficult to detect than a simple cutoff. The high-energy emission from residual collisions is also expected to be delayed relative to MeV  $\gamma$ -rays. It should be noticed that we only focus on the *prompt* high-energy emission which appears simultaneously with the MeV  $\gamma$ -ray emission.

We show first in Section 2 the strong  $\gamma\gamma$  absorption during the MeV  $\gamma$ -ray emitting phase, next discuss in Section 3 the high-energy emission from residual collisions, then the model is applied to the recent *Fermi* detection of GRB 080916c (Abdo et al. 2009) in Section 4, which might have provided evidences of the model, and finally, a general discussion on observations is given in Section 5.

## 2. $\gamma\gamma$ ABSORPTION AT SMALL RADII

Consider a highly relativistic outflow with a bulk Lorentz factor  $\Gamma = 10^{2.5}\Gamma_{2.5}$ . The small timescale variation will lead to strong internal shocks, which produce  $\gamma$ -ray emission. Let us denote the radius where  $\gamma$ -rays arise by  $R_\gamma$ . Due to geometry effect, the observed fastest variability timescale  $t_{\text{var}} \lesssim 10^{-2}$  s (e.g., Woods & Loeb 1995) in  $\gamma$ -ray light curves suggests that the size of  $\gamma$ -ray emission region is limited to  $R_\gamma \lesssim 2\Gamma^2 ct_{\text{var}}$ . If the  $\gamma$ -ray emitting electrons are fast cooling (with cooling time shorter than the dynamical time of the outflow), we should take the equality, thus,  $R_\gamma \approx 6 \times 10^{13} L_{2.5}^2 t_{\text{var},-2}^2$  cm, where  $t_{\text{var}} = 10^{-2} t_{\text{var},-2}$  s. Actually, in the context of internal shock models, we do not expect that the magnetic field is generated with energy density much higher than that of accelerated electrons. In order to have synchrotron emission peaking at  $\epsilon_b \sim 1$  MeV, as observed in GRBs, the radius of the  $\gamma$ -ray emitting region should be small (LW08),  $R_\gamma \lesssim 10^{13} L_{\text{bol},52}^{1/2} (\epsilon_b/1 \text{ MeV})^{-1}$  cm, where  $L_{\text{bol}} = 10^{52} L_{\text{bol},52} \text{ erg s}^{-1}$  is the bolometric  $\gamma$ -ray luminosity.

Let us consider the  $\gamma\gamma$  absorption due to pair production inside the GRB source. For a photon of high-energy  $\epsilon$  the optical depth to pair production within the GRB source is given by the product of the pair-production rate,  $1/t'_{\gamma\gamma}(\epsilon)$ , and the dynamical time, the time required for significant expansion of the plasma,  $t'_d \simeq R_\gamma/\Gamma c$  (primes denote quantities measured in the outflow rest frame),  $\tau_{\gamma\gamma}(\epsilon) \simeq R_\gamma/\Gamma c t'_{\gamma\gamma}(\epsilon)$ . Here,  $t'_{\gamma\gamma}(\epsilon)$  depends on the energy density and on the spectrum of the radiation. The (outflow rest frame) radiation energy density is approximately given by  $U'_\gamma = L/4\pi R_\gamma^2 c \Gamma^2$ . The GRB spectrum can typically be described as a broken power law,  $dn/d\epsilon \propto \epsilon^{-\beta}$ , with  $\beta \approx -1$  at low energy,  $\epsilon < \epsilon_b \sim 1$  MeV, and  $\beta \approx -2$  at  $\epsilon > \epsilon_b$  (Band et al. 1993). High-energy photons with energy  $\epsilon'$  exceeding  $\epsilon'_b \equiv 2(m_e c^2)^2/\epsilon'_b$  may produce pairs in interactions with photons of energy exceeding  $\epsilon' = 2(m_e c^2)^2/\epsilon' < \epsilon'_b$  (the rest frame photon energy  $\epsilon'$  is related to the observed energy by  $\epsilon = \Gamma\epsilon'$ ). For  $\epsilon' < \epsilon'_b$  we have  $dn/d\epsilon' \propto \epsilon'^{-1}$ , which implies that the number density of photons with energy exceeding  $\epsilon'$  depends only weakly on energy. Thus,  $t_{\gamma\gamma}$  is nearly independent of energy for  $\epsilon' > \epsilon'_b$ ,  $t'_{\gamma\gamma} \approx (\sigma_T/16) c U'_\gamma / \epsilon'_b$ , which gives

$$\tau_{\gamma\gamma}(\epsilon > \epsilon_b) \simeq 1.1 \times 10^2 \frac{L_{52}}{t_{\text{var},-2} \Gamma_{2.5}^4} \left( \frac{\epsilon_b}{1 \text{ MeV}} \right)^{-1}. \quad (1)$$

Note, we have approximated the  $\gamma\gamma$  cross section above the pair-production threshold as  $3\sigma_T/16$  (Waxman 2003). Also note that as the energy density  $U'_\gamma$  (and hence the related luminosity  $L$ ) depends on the energy band considered, the one used in calculating  $t'_{\gamma\gamma}$  is the energy density below  $2 \times \epsilon'_b$ . Hereafter, without special emphasis the luminosity  $L = 10^{52} L_{52} \text{ erg s}^{-1}$  is the so-called MeV luminosity, only corresponding to emission at  $< 2 \times \epsilon_b$ , i.e.,  $L \equiv \int_0^{2\epsilon_b} L_\epsilon d\epsilon$ . Photons of lower energy,  $\epsilon < \epsilon_b$ , interact to produce pairs only with photons of energy  $\epsilon' > 2(m_e c^2)^2/\epsilon' > \epsilon'_b$ . Since the number density of these photons drops like  $1/\epsilon'$ , we have  $\tau_{\gamma\gamma}(\epsilon < \epsilon_b) \approx (\epsilon/\epsilon_b) \tau_{\gamma\gamma}$

( $\epsilon > \epsilon_b$ ), i.e.,

$$\tau_{\gamma\gamma}(\epsilon < \epsilon_b) \simeq 2.2 \times 10^{-3} \frac{L_{52}}{t_{\text{var},-2} \Gamma_{2.5}^6} \frac{\epsilon}{1 \text{ MeV}}. \quad (2)$$

The optical depth increases as photon energy increases. Photons with high enough energy are absorbed in the emission region. A spectral cutoff is defined by  $\tau_{\gamma\gamma} = 1$ ,<sup>3</sup>

$$\epsilon_{\text{cut}}^{(1)} \simeq 0.46 \frac{t_{\text{var},-2} \Gamma_{2.5}^6}{L_{52}} \text{ GeV}. \quad (3)$$

A comment on the approximation of the  $\gamma\gamma$  cross section should be made here. Since both the cross section and the GRB photon spectrum decrease rapidly with photon energy, the approximation is excellent—for a GRB spectrum with  $\beta \approx -2$ , it precisely produces the optical depth within 2%, compared to a calculation with full cross section.

There is another restriction for the cutoff besides Equation (3). In deriving the cutoff, Equation (3), the optical depth is not self-consistently calculated since a GRB spectrum extending to infinity without a high-energy cutoff is assumed. Given the two factors that GRB spectrum usually appears to be a steep slope, with the photon number dominated by low energy photons, and that the high ( $\epsilon' > 2^{1/2} m_e c^2$ ) and low ( $\epsilon' < 2^{1/2} m_e c^2$ ) energy photons annihilate each other one by one, we only expect high-energy photons might be totally attenuated by low energy ones, other than the other way around. So the cutoff should be  $\epsilon'_{\text{cut}} > 2^{1/2} m_e c^2$ , which is not assured by  $\epsilon_{\text{cut}}^{(1)}$  in Equation (3) ( $\epsilon_{\text{cut}}^{(1)}/\Gamma < 2^{1/2} m_e c^2$  might happen). We need to define, in the GRB source frame,

$$\epsilon_{\text{cut}}^{(2)} = 2^{1/2} \Gamma m_e c^2 \simeq 0.23 \Gamma_{2.5} \text{ GeV}. \quad (4)$$

Note, this condition is also forgotten by many other authors who calculate the cutoff energy assuming a no-cutoff GRB spectrum. The GRB spectral cutoff is, therefore, expected at the maximum between  $\epsilon_{\text{cut}}^{(1)}$  and  $\epsilon_{\text{cut}}^{(2)}$  (Li & Song 2004),

$$\epsilon_0 \equiv \epsilon_{\text{cut}}(R_\gamma) \simeq \max \left[ 0.46 \frac{t_{\text{var},-2} \Gamma_{2.5}^6}{L_{52}}, 0.23 \Gamma_{2.5} \right] \text{ GeV}. \quad (5)$$

A critical Lorentz factor where  $\epsilon_{\text{cut}}^{(1)} = \epsilon_{\text{cut}}^{(2)}$  is a function of  $R_\gamma$  (hence  $t_{\text{var}}$ ),

$$\Gamma_c \simeq 280 (L_{52}/t_{\text{var},-2})^{1/5}. \quad (6)$$

$\epsilon_0$  sensitively depends on  $\Gamma$  thus detection of the spectral cutoff may be very useful in constraining  $\Gamma$ ,

$$\Gamma = \min \left[ 360 \left( \frac{L_{52}}{t_{\text{var},-2}^{\text{ob}} \Gamma_{2.5}^6} \frac{\epsilon_0^{\text{ob}}}{1 \text{ GeV}} \right)^{1/6} (1+z)^{1/3}, \right. \\ \left. 1390 \frac{\epsilon_0^{\text{ob}}}{1 \text{ GeV}} (1+z) \right]. \quad (7)$$

Here the redshift factor has been included, i.e.,  $t^{\text{ob}} = t(1+z)$  and  $\epsilon^{\text{ob}} = \epsilon/(1+z)$ .

The extension of GRB spectra to  $\gtrsim 100$  MeV and the characteristic variability time,  $t_{\text{var},-2} \lesssim 1$ , have implied  $\Gamma_{2.5} \gtrsim 1$ ,

<sup>3</sup> Because a low energy turnover at  $\epsilon_a \sim 1$  keV is expected in GRB spectra due to synchrotron self-absorption, very high energy photons with  $\epsilon \gtrsim 10^{16} L_{52} t_{\text{var},-2}^{-1} \Gamma_{2.5}^{-2} (\epsilon_a/1 \text{ keV})^{-1} \text{ eV}$  still can escape from the GRB source (e.g., Li & Waxman 2007; see also Razzaque et al. 2004).

assuming the  $\gtrsim 100$  MeV photons are produced in the same time and place as the MeV photons. Since thermal pressure acceleration in the initial fireball cannot lead to much larger Lorentz factors, and in the internal shock model much larger Lorentz factors would lead to synchrotron emission peaking below MeV band (see discussion in the first paragraph of Section 2),  $\Gamma_{2.5} \approx 1$  is typically adopted (e.g., Waxman 2003). The exact values of  $\Gamma$  would be determined by the detection of the high-energy cutoffs in GRB spectra by, e.g., *Fermi*. We show below that the situation may be complicated by the delayed large-radius emission from the outflow.

### 3. LARGE-RADIUS EMISSION FROM RESIDUAL COLLISIONS

After the initial strong collisions at small radii driven by the small timescale,  $\sim 1$  ms, variation in the outflow, there are residual collisions continue to occur when the outflow is expanding to large radii. As the velocity fluctuation in the flow is being smoothed out by the on-going collisions, the delayed collisions become weaker, and the postshock electron energy and magnetic field are smaller, which give rise to synchrotron emission at longer wavelengths. LW08 has well discussed the dynamics of the residual collisions and naturally explained the optical flashes from GRBs by this delayed residual emission. As mentioned in LW08, the energy density in the delayed collisions is dominated by the primary  $\gamma$ -ray emission, so that the residual emission is dominated by IC scattering of the prompt  $\gamma$ -ray photons. In what follows consider the IC emission.

#### 3.1. Dynamics

Let us approximate the outflow by a sequence of  $N \gg 1$  individual shells. LW08 has considered the simplified case: the shell masses are equal; the shells are initially separated by a fixed distance  $ct_{\text{var}}$ ; the extent of the shells is much smaller than the radius of the outflow; and the Lorentz factors of individual shells are drawn from a random distribution with an average  $\Gamma$  and initial variance  $\sigma_{\Gamma,0}^2 < \Gamma^2$ . The model may be more complicated, by adding more degrees of freedom, however, LW08 has shown that this simple case naturally accounts for the observed properties of GRB optical flashes. Here we will consider this simplified case. Moreover, we adopt the assumption that two shells merge into a bigger shell after a collision, i.e., the full inelastic collision where the internal energy generated is fully radiated. If the postshock electrons carry an energy close to equipartition and the electrons cool fast, the internal energy is always radiated significantly. In the case of a significant fraction of the internal energy being dissipated in each collision, the dynamical evolution of the outflow has been proven to be similar to the full inelastic case (LW08).

In the simple case under consideration, the variance of the velocities of individual shells (in the outflow rest frame) evolves with the outflow radius  $R$  as  $\sigma_v \propto R^{-1/3}$ . The outflow energy that is associated with the fluctuation of shell velocities (in the outflow rest frame) and hence may be dissipated decreases as

$$E_{\text{fluc}} \propto \Gamma \sigma_v^2 \propto R^{-2/3}. \quad (8)$$

In general, it is naturally expected that there might be a wide range of variability timescale,  $\sim 1$  ms–10 s, in the flow properties. Large timescale variabilities might lead to more energy dissipated at large radii. Thus, the slope should be flatter than  $-2/3$ . If a power-law description,  $E_{\text{fluc}} \propto R^{-q}$ , is still

available, one may have  $0 < q < 2/3$ . We carry a Monte Carlo simulation to demonstrate this point in Appendix A.

#### 3.2. High-energy Emission

Based on the dynamical evolution, the emission from the residual collisions can be further predicted. Taking the common assumptions that in internal shocks the postshock electrons and magnetic fields carry fixed fractions,  $\epsilon_e$  and  $\epsilon_B$ , respectively, of the postshock internal energy, the characteristic Lorentz factor of postshock electrons (in the outflow rest frame) scales as  $\gamma_i \propto \epsilon_e \sigma_v^2 \propto R^{-2/3}$ , and the postshock magnetic field scales as  $B^2 \propto \epsilon_B \sigma_v^2 n_e \propto R^{-8/3}$  (the particle number density scales as  $n_e \propto R^{-2}$ ).

We demonstrate that the electrons in residual collisions lose most of their energy by IC cooling. If the prompt  $\gamma$ -rays last a duration  $T$  (observer frame), the plasma is overlapped with these  $\gamma$ -rays until the outflow expands to  $R \gtrsim 2\Gamma^2 cT \simeq (T/t_{\text{var}})R_\gamma$ . LW08 showed that when the synchrotron emission lies in the optical band, the radius is  $R_{\text{opt}} \simeq 10^2 R_\gamma$ . For typical observed values  $t_{\text{var}} \lesssim 10^{-2}$  s and  $T \sim 10$  s, the optical radius is still relatively small,  $R_{\text{opt}} < 2\Gamma^2 cT$ . Therefore, during the phase of late residual collisions that we concern, the plasma is immersed in the radiation bath of the prompt  $\gamma$ -rays. Both the photon energy density  $U_\gamma$  and the particle number density  $n_e$  drop as  $\propto R^{-2}$  hence the ratio  $y = U_\gamma/(B^2/8\pi) \propto \sigma_v^{-2} \propto R^{2/3}$  increases with  $R$ . Because  $y \sim 1$  in the  $\gamma$ -ray producing phase, we have  $y > 1$  in residual collision phase, so the radiation energy density dominates that of the magnetic field. Let us consider the properties of IC emission.

##### 3.2.1. Spectrum

Consider first the energy band into which energy is radiated. The prompt  $\gamma$ -ray photons with typical energy  $\epsilon_b$  are up-scattered by electrons with characteristic Lorentz factor  $\gamma_i$  to energy  $\epsilon_{\text{IC}} \simeq \lambda \gamma_i^2 \epsilon_b \simeq \lambda \epsilon_e^2 (m_p/m_e)^2 (R/R_\gamma)^{-4/3} \epsilon_b$ . Here we assume  $\gamma_{i,0} \sim \epsilon_e (m_p/m_e)$  as the electron Lorentz factor emitting MeV  $\gamma$ -rays, and  $\lambda$  accounts for the correction due to uncertain geometry effect. It will be shown in Appendix B that the correction factor  $\lambda$  is order unity even in the case that the prompt MeV photons are highly beamed in the rest frame of the outflow. The characteristic scattered photon energy is

$$\epsilon_{\text{IC}}^{\text{ob}} \simeq 9\lambda \epsilon_e^2 \frac{\epsilon_b}{1 \text{ MeV}} \left( \frac{R}{10^2 R_\gamma} \right)^{-4/3} (1+z)^{-1} \text{ GeV}. \quad (9)$$

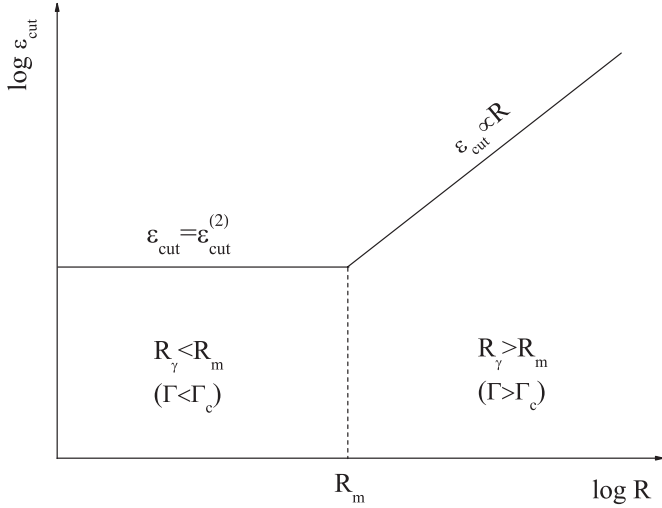
The scattering might take place within slight Klein–Nishina regime,  $\gamma_i \epsilon_b'/m_e c^2 \sim$  a few  $> 1$ , at small radii  $R \sim R_\gamma$ , where the energy of scattered photons is instead  $\epsilon_{\text{IC}} = \Gamma \gamma_i m_e c^2 = \epsilon_e \Gamma (R/R_\gamma)^{-2/3} m_p c^2 \simeq 3 \times 10^2 \epsilon_e (R/R_\gamma)^{-2/3} \text{ GeV}$ .

Next consider the  $\gamma\gamma$  absorption effect on the late residual emission. For GRB outflow with  $\Gamma > \Gamma_c$ , the initial cutoff energy for the primary emission is determined by the first term in the bracket of Equation (5),  $\epsilon_0 = \epsilon_{\text{cut}}^{(1)}$ . Equation (3) implies that the spectral cutoff energy scales as  $\epsilon_{\text{cut}} \propto R \Gamma^4 L^{-1}$ , so the cutoff energy increases with  $R$  for fixed  $L$ . We have, for late residual collisions,  $\epsilon_{\text{cut}} \simeq \epsilon_0 R/R_\gamma$ . In the case of GRB outflow with lower Lorentz factor  $\Gamma < \Gamma_c$ , the cutoff energy is initially a constant,  $\epsilon_0 = \epsilon_{\text{cut}}^{(2)}$  (Equation 5), until the outflow expands to a radius,

$$R_m = 3 \times 10^{13} L_{52} \Gamma_{2.5}^{-3} \text{ cm} \quad (10)$$

(note  $R_m > R_\gamma$ ). At  $R > R_m$ , the cutoff energy turns to increase with  $R$ ,  $\epsilon_{\text{cut}} \simeq \epsilon_0 R/R_m$ . In both cases of  $\Gamma > \Gamma_c$  and  $\Gamma < \Gamma_c$  the





**Figure 1.** Schematic plot of the attenuated energy evolving with radius. There are two regimes. If the primary collisions that produce MeV emission occur in the regime of  $R_\gamma > R_m$  (i.e.,  $\Gamma > \Gamma_c$ ),  $\varepsilon_0 = \varepsilon_{\text{cut}}^{(1)}$  and  $\varepsilon_{\text{cut}} \propto R$  later. If  $R_\gamma < R_m$  ( $\Gamma < \Gamma_c$ ), in the beginning the attenuated energy is a constant,  $\varepsilon_{\text{cut}} = \varepsilon_{\text{cut}}^{(2)}$  at  $R < R_m$ , and turns to be  $\varepsilon_{\text{cut}} \propto R$  at  $R > R_m$ .

cutoff energy at  $R > \max[R_\gamma, R_m]$  (i.e.,  $\varepsilon_{\text{cut}} > \varepsilon_0$ ) follows the same expression,

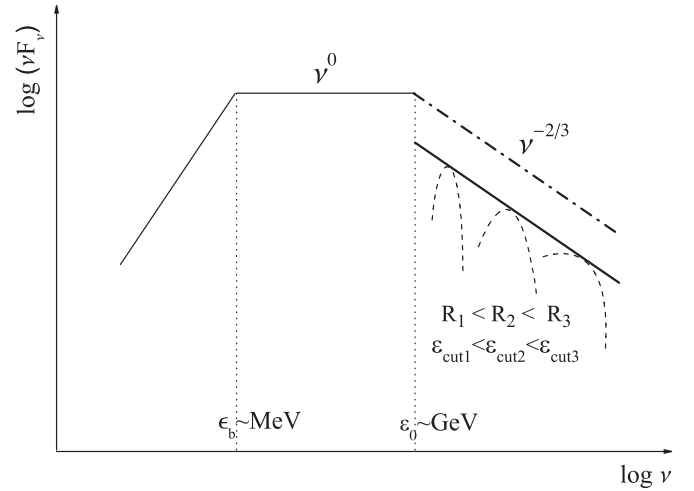
$$\varepsilon_{\text{cut}}^{\text{ob}} \simeq 50 \frac{t_{\text{var}, -2}^{\text{ob}} \Gamma_{2.5}^6}{L_{52}} \left( \frac{R}{10^2 R_\gamma} \right) (1+z)^{-2} \text{ GeV}. \quad (11)$$

The evolution of  $\varepsilon_{\text{cut}}$  versus  $R$  for fixed  $L$  is plotted in Figure 1.

Comparing  $\varepsilon_{\text{IC}}$  and  $\varepsilon_{\text{cut}}$  it can be found that typically  $\varepsilon_{\text{IC}} > \varepsilon_{\text{cut}}$  at small radii,  $R \lesssim 30 R_\gamma$ . In this case, the bulk IC radiation is absorbed in the source, leading to electromagnetic cascades, and the photons escape until their energies decay to  $\varepsilon_{\text{cut}}$ . Therefore, the bulk high-energy radiation is just re-emitted at  $\varepsilon_{\text{cut}}$ . On the other hand,  $\varepsilon_{\text{IC}} < \varepsilon_{\text{cut}}$  at large radii  $R \gtrsim 30 R_\gamma$ , where only the high-energy emission below  $\varepsilon_{\text{cut}}$  appears. The emission above  $\varepsilon_{\text{cut}}$  is truncated, and undergoes electromagnetic cascades, but does not affect much the apparent spectrum since the photon spectrum decreases rapidly with energy.

Finally consider the emission flux. It is straightforward to show that the cooling time of the electrons is short compared to the dynamical time during the late residual collision phase, up to radii  $R \sim 10^3 R_\gamma$ . We therefore assume that electrons radiate away all their energy. When  $\varepsilon_{\text{IC}} > \varepsilon_{\text{cut}}$  at small radii, the total electron energy appears at  $\varepsilon_{\text{cut}} \simeq \varepsilon_0$  if  $R_\gamma < R < R_m$  or  $\varepsilon_{\text{cut}} \propto R$  if  $R > \max[R_\gamma, R_m]$ . The observed (time-integrated) IC spectrum at energy  $\varepsilon > \varepsilon_0$  would be  $\nu F_\nu \propto E_{\text{fluc}}|_{\nu_{\text{cut}}=\nu} \propto R^{-2/3}|_{\nu_{\text{cut}}=\nu} \propto \nu^{-2/3}$  (here  $\nu = \varepsilon/h$ ).

When the outflow expands to large radii where  $\varepsilon_{\text{IC}} < \varepsilon_{\text{cut}}$ , we need to consider the electrons accelerated to Lorentz factors larger than the characteristic Lorentz factor  $\gamma_i$ . Shock acceleration is expected to generate a power-law energy distribution of electrons  $dn_e/d\gamma_e \propto \gamma_e^{-p}$  at  $\gamma_e > \gamma_i$  with  $p \simeq 2$  (Blandford & Eichler 1987). This flat-electron energy distribution,  $\gamma_e^2 dn_e/d\gamma_e \propto \gamma_e^0$ , generates equal amounts of IC energy in logarithmic photon energy intervals,  $\nu F_\nu \propto \nu^0$  for  $\nu > \nu_{\text{IC}}$ . So  $\nu F_\nu(\nu_{\text{cut}}) \simeq \nu F_\nu(\nu_{\text{IC}}) \propto E_{\text{fluc}}$ . The emission at low energy would be covered by earlier emission, while only the emission at high-energy end, i.e., around the cutoff, shows up and interests us. The observed (time-integrated) spectrum would be similar to the  $\varepsilon_{\text{IC}} > \varepsilon_{\text{cut}}$  case, i.e.,  $\nu F_\nu \propto \nu^{-2/3}$ .



**Figure 2.** Schematic plot of the predicted  $\nu F_\nu$  spectrum of prompt high-energy emission from a GRB. The thin solid line shows the observed MeV  $\gamma$ -ray emission, a broken power law with a break energy at  $\varepsilon_b \sim \text{MeV}$ , above which the spectrum goes as  $\nu F_\nu \propto \nu^0$ . The dotted lines mark the break energy  $\varepsilon_b$  and the previously thought pair-production spectral cutoff,  $\varepsilon_0 \sim \text{GeV}$  (Equation (5)). The residual collisions at large radii contribute beyond  $\varepsilon_0$ . Summing up all emission components (the dashed lines) from different radii and times lead to a spectral slope  $\nu F_\nu \propto \nu^{-q}$  ( $q$  is the index of the random energy evolution,  $E_{\text{fluc}} \propto R^{-q}$ , and  $0 < q \leq 2/3$ ). In the simplified case,  $q = 2/3$  (see the text). Here the thick solid line corresponds to the outflow satisfying  $R_\gamma < R_m$  (or  $\varepsilon_0 \simeq \Gamma m_e c^2$ ), while the thick dash-dotted line corresponds to  $R_\gamma > R_m$  ( $\varepsilon_0 \gtrsim \Gamma m_e c^2$ ). The  $\nu F_\nu$  values in these two are different by a factor of  $(R_m/R_\gamma)^{2/3}$ . The spectral slope holds up to TeV range, but not higher (see Section 5).

Thus, we expect the observed (time-integrated) prompt spectrum, above the spectral cutoff energy in the prompt  $\gamma$ -ray emitting phase,  $\varepsilon_0$  (Equation (5)), to be

$$\nu F_\nu \propto \nu^{-2/3}, \quad h\nu > \varepsilon_0 \text{ (simplified case)}. \quad (12)$$

This fluence spectrum is resulted from summing up all emission components from different radii and times. A schematic plot of the prompt GRB spectrum is shown in Figure 2. Note, the spectral slope  $\nu^{-2/3}$  here is derived from the simple case, which has been confirmed by recent numerical calculation by Aoi et al. (2009). If in general  $E_{\text{fluc}} \propto R^{-q}$  we would expect the slope to be  $\nu F_\nu \propto \nu^{-q}$ .

Below  $\varepsilon_0$  is the observed prompt MeV  $\gamma$ -ray spectrum, i.e., typically  $\nu F_\nu \propto \nu$  below  $\varepsilon_b$  and  $\nu F_\nu \propto \nu^0$  between  $\varepsilon_b$  and  $\varepsilon_0$ . Note, the transition of the emission from primary to residual collisions at  $\varepsilon_0$  is smooth if  $R_\gamma > R_m$ , as shown by the dash-dotted line. The transition for the case of  $R_\gamma < R_m$  is discontinuous as shown by the thick solid line. The power law described by Equation (12) starts with a flux lower than the primary emission at  $\varepsilon_0$  by a factor of  $(R_m/R_\gamma)^{2/3}$ .

Some comments should be made here. The spectral form described in Equation (12) holds only on average, especially for the high-energy range. In individual GRB events the flux may differ significantly, because, for a small number of shells (and collisions) at large radii, large variations in the late residual collisions should be expected.

It should also be noticed that we have assumed the initial variance  $\sigma_{\Gamma,0} < \Gamma$ , whereas initial condition with  $\sigma_{\Gamma,0} > \Gamma$  may lead to more efficient  $\gamma$ -ray production (e.g., Beloborodov 2000) around  $R_\gamma$ , in which case the ratio between fluxes of primary and residual emission at  $\varepsilon_0$  should be larger by a factor

of a few,<sup>4</sup> leading to more abrupt transition between the primary and residual emission.

One may worry about that the IC emission may be reduced as the seed photons are not isotropic in the shock frame of residual collisions. However, this kind of geometry effects does not play an important role even if the photons are completely collinear (Wang et al. 2006). As usually assumed by many authors, suppose that the electrons accelerated in residual collision shocks are isotropic in the rest frame of the outflow, since the tangled magnetic fields in the shock might sufficiently isotropize electrons. Thus, an electron changes its angle  $\theta'$  with respect to the photon beam and cools fast. The IC power of an electron averaged over its cooling time is not different from interacting with isotropic photons. As long as the jet effect is not important to prompt GRB emission ( $\theta_{\text{jet}} \gg 1/\Gamma$ ), we can furthermore regard the GRB explosion as isotropic, and hence the observer at different angles will observe the same IC emission due to spherical symmetry. Consider both cases of isotropic and anisotropic scatterings, the radiated energy can be assumed to be the same because it is determined by the total electron energy if electrons radiate all their energy rapidly. The outside observers would observe the same fluence in both cases, otherwise, one can simply ask where the electron energy have gone, given the same total electron energy. Thus, the IC fluence is not reduced by this geometry effect if electrons are isotropic distributed in the rest frame and radiate all their energy within a dynamic time. Nevertheless, this effect changes the angular distribution of IC emission. In the rest frame the IC power becomes  $P'_{\text{IC}} \propto (1 - \cos \theta')^2$ , although not much different from isotropic distribution. Correspondingly, in the lab frame the “image” of the IC emission is different from the isotropic case, i.e., the angular dependence of the brightness is different, but the angular-integrated fluence is the same.

### 3.2.2. Time Delay

At energy  $\varepsilon < \varepsilon_0$ , the emission is mainly contributed by primary collisions at small radius and arrives at detectors simultaneously with MeV emission. However, for higher energy  $\varepsilon > \varepsilon_0$ , the emission is produced at relatively large radii and should have a time delay relative to the primary MeV emission. Since the spectral cutoff energy, where the high-energy photons emerge, increases as the outflow expands,  $\varepsilon_{\text{cut}} \propto R$  for  $\varepsilon_{\text{cut}} > \varepsilon_0$  (Equation 11), the time delay increases with observed photon energy. For energy  $\varepsilon = \varepsilon_{\text{cut}}(R) > \varepsilon_0$ , the typical radius where photons emerge is  $R \simeq [\varepsilon/\varepsilon_0] \max[R_\gamma, R_m] = [\varepsilon/\varepsilon_{\text{cut}}^{(1)}] R_\gamma$ , therefore the related time delay relative to MeV  $\gamma$ -ray emission,  $\tau_{\text{delay}} \simeq (R - R_\gamma)/2\Gamma^2 c$ , is

$$\begin{aligned} \tau_{\text{delay}}^{\text{ob}} &\simeq \frac{\varepsilon}{\varepsilon_{\text{cut}}^{(1)}} t_{\text{var}} \\ &= 2.2 L_{52} \Gamma_{2.5}^{-6} \frac{\varepsilon^{\text{ob}}}{10^2 \text{ GeV}} (1+z)^2 \text{ s (for } \varepsilon^{\text{ob}} > \varepsilon_0^{\text{ob}}). \end{aligned} \quad (13)$$

Thus, the emission at  $\varepsilon > \varepsilon_0$  is delayed later as  $\varepsilon$  increases.

A comment is made here that the target photons for scatterings may be beamed with respect to electrons, which changes the angular distribution of IC emission, i.e., the maximum IC power may come from a certain angle other than  $\theta = 0$ .

This leads to an additional time delay. However, as implied by Equation (B5), the angle where the maximum power is emitted is smaller than  $\pi/2$  in the rest frame and hence  $\theta < 1/\Gamma$ . The produced time delay is smaller than  $\tau_{\text{delay}}$ ,  $R\theta^2/2c < R/2\Gamma^2 c \sim \tau_{\text{delay}}$ . Thus, we neglect the additional contribution of time delay by this anisotropic-scattering effect. Moreover, the additional time delay does not smear out the light-curve variability much more than the usual used spreading time, thus, we still have the variability time in high-energy emission,  $t_{\text{var,HE}} \sim R/\Gamma^2 c \sim \tau_{\text{delay}}$ .

Equation (13) implies that the detection of time delay  $\tau_{\text{delay}}$  at  $\varepsilon$  helps to determine the Lorentz factor,

$$\Gamma = 167 L_{52}^{1/6} \left( \frac{\varepsilon^{\text{ob}}}{1 \text{ GeV}} \right)^{1/6} \left( \frac{\tau_{\text{delay}}^{\text{ob}} |_{\varepsilon^{\text{ob}}}}{1 \text{ s}} \right)^{-1/6} (1+z)^{1/3}. \quad (14)$$

This determination by time delay should be consistent with that by detection of the spectral transition  $\varepsilon_0$  between primary and residual emission, Equation (7).

## 4. APPLICATION: FERMI-LAT GRB 080916c

As the *Fermi* observational data showed up after the first version of this paper was posted on the archive,<sup>5</sup> we now apply the model to the observations.

Abdo et al. (2009) report the measurements of the bright GRB 080916c by *Fermi* GBM and LAT. The redshift of this burst is  $z = 4.35$  (Greiner et al. 2009b), which implies, with fluence (10 keV–10 GeV)  $\approx 2.4 \times 10^{-4} \text{ erg cm}^{-2}$ , the largest reported isotropic  $\gamma$ -ray energy release,  $E_{\text{iso}} \simeq 9 \times 10^{54} \text{ erg}$ . The observed GRB duration is  $T \approx 50 \text{ s}$ , so the bolometric isotropic-equivalent luminosity is  $L_{\text{bol}} = E_{\text{iso}}(1+z)/T \approx 10^{54} \text{ erg s}^{-1}$ . As the observed peak energy is  $\varepsilon_b/(1+z) \sim 1 \text{ MeV}$  and the high-energy slope is  $\beta \approx -2$ , the MeV luminosity, defined as the luminosity at  $\lesssim 2\varepsilon_b$ , is  $L \approx L_{\text{bol}}/\ln(10 \text{ GeV}/1 \text{ MeV}) = 10^{53} \text{ erg s}^{-1}$ . The LAT detected 145 photons at  $>100 \text{ MeV}$ , within which 14 are beyond 1 GeV, during the first 100 s after the trigger. The brightness may have enough statistics for spectral and temporal analysis of the high-energy properties.

There are several interesting properties in the high-energy emission of this GRB.

*Time delay.* The multiband light curves unambiguously show that the bulk of the emission of the second light-curve peak is moving toward later times as the energy increases (see time bin “b” in Figure 1 and its inset panels in Abdo et al. 2009), and the time delay of 100 MeV emission is about 1 s relative to MeV emission, much larger than the MeV variability timescale,  $\lesssim 100 \text{ ms}$  (Greiner et al. 2009b).<sup>6</sup> First of all, these are qualitatively consistent with our prediction that the higher energy photons can only arise when the plasma expands to larger size in later time where the  $\gamma\gamma$  optical depth reduces to below unity, and that the size of high-energy emission can be much larger than MeV emission.

Let us consider the data quantitatively and constrain the model parameters. The LAT  $>100 \text{ MeV}$  detection consists of 145 photons which mainly come up in a single light-curve peak, therefore we have enough statistics for the time analysis of

<sup>5</sup> <http://arxiv.org/abs/0810.2932>

<sup>6</sup> Note, the time-delay issue here is different from what is called “delayed onset” by other authors. We concern indeed the delayed peaking time of high-energy emission, related to the delayed arrival of the bulk of high-energy emission.

<sup>4</sup> It is not expected that the initial variance of Lorentz factors is far exceeding the mean,  $\sigma_{\Gamma,0} \gg \Gamma$ .

>100 MeV emission. It is obviously seen that the >100 MeV light curve peak has a time delay  $\tau_{\text{delay}}^{\text{ob}} = \tau_{\text{delay}}(1+z) \sim 1$  s relative to that of 250 keV–5 MeV (the superscript “ob” denotes quantities measured in the observer frame, with redshift effect taken into account). This implies  $\varepsilon_0^{\text{ob}} = \varepsilon_0/(1+z) < 100$  MeV. Substituting the observed values of  $L = 10^{53}$  erg s $^{-1}$ ,  $\tau_{\text{delay}} \simeq 1/5.35$  s, and  $\varepsilon = 100 \times 5.35 = 535$  MeV, with redshift  $z = 4.35$  taken, into Equation (13), we obtain the bulk Lorentz factor of GRB 080916c outflow,

$$\Gamma \simeq 290 \left( \frac{L}{10^{53} \text{ erg s}^{-1}} \right)^{1/6} \left( \frac{\tau_{\text{delay}}^{\text{ob}}|_{100 \text{ MeV}}}{1 \text{ s}} \right)^{-1/6}. \quad (15)$$

This result is similar to those determined in other GRBs,  $\Gamma \approx 100$ –400, through observations of the rising of optical afterglows (Molinari et al. 2007; Krühler et al. 2008, 2009; Greiner et al. 2009a), and the thermal components in the prompt emission (Pe’er et al. 2007).

The determination of  $\Gamma$  can be double checked by the location of  $\varepsilon_0$ . Using the result of Equation (15), we obtain

$$\varepsilon_{\text{cut}}^{(1)\text{obs}} \simeq 10(t_{\text{var}}^{\text{ob}}/10^2 \text{ ms})(\tau_{\text{delay}}^{\text{ob}}|_{100 \text{ MeV}}/1 \text{ s})^{-1} \text{ MeV}$$

and

$$\varepsilon_{\text{cut}}^{(2)\text{ob}} \simeq 40(L/10^{53} \text{ erg s}^{-1})^{1/6}(\tau_{\text{delay}}^{\text{ob}}|_{100 \text{ MeV}}/1 \text{ s})^{-1/6} \text{ MeV}.$$

The observed MeV variability timescale is  $t_{\text{var}}^{\text{ob}} \lesssim 100$  ms based on the *INTEGRAL* observation (Greiner et al. 2009b). Thus  $\varepsilon_0^{\text{ob}} \sim 40$  MeV, consistent with requirement  $\varepsilon_0^{\text{ob}} < 100$  MeV. In addition, the broad light-curve peak in the no-energy-selection band of LAT is consistent with, or a little delayed from, that of the GBM (260 keV–5 MeV) light curve, and is ahead of the >100 MeV peak. Thus  $\varepsilon_0^{\text{ob}}$  should be located in the LAT energy window (no selection) and below 100 MeV, consistent with the result  $\varepsilon_0^{\text{ob}} \sim 40$  MeV.

By our model the >1 GeV emission should be delayed even 10 times longer than the >100 MeV one, i.e.,  $\tau_{\text{delay}}^{\text{ob}}|_{1 \text{ GeV}} \sim 10$  s. The much fewer photons above 1 GeV prevent us from analyzing the temporal properties with high confidence. However the LAT >1 GeV light curve does agree with a longer delay by  $\sim 10$  s.

It should be noticed that other authors also constrain the bulk Lorentz factor of this GRB and obtain much larger lower limit (Greiner et al. 2009b; Abdo et al. 2009). Essentially, the difference is due to different models considered; they consider the GeV emission produced in the same time and place as the MeV emission, whereas in our model the GeV emission comes from delayed residual collisions at large radii, therefore our model loses the constraint on  $\Gamma$ . In addition, we consider that the cutoff energy should not locate below  $2^{1/2}\Gamma m_e c^2$ , which is ignored in Greiner et al. (2009b) and Abdo et al. (2009).

*Lack of the first LAT light-curve peak.* The low energy GBM light curve shows two peaks, however, the LAT observations only show one peak related to the second GBM peak and there is a paucity of emission in the first  $\sim 4$  s after the trigger. Note, some people call this as a “delayed onset” of high-energy emission. We do not try to explain this feature in details. However, in principle one of the explanations could be the residual-collision model.

The *Fermi* results show that a Band function fit gives a high-energy index consistent with  $\beta \approx -2/3$ . This can be explained by residual-collision emission from an outflow with single-

timescale variability, and the initial cutoff happens at few MeV. On the other hand, the spectrum of the first GBM light-curve peak might also be fit with a cutoff at  $\sim 10$  MeV. If so, we can use specific parameters of the early 4 s ejecta, i.e., bulk Lorentz factor and variability timescales, in order that the initial spectral cutoff is  $\sim 10$  MeV and the residual-collision emission at >10 MeV is  $\gtrsim 4$  s delayed in observations, longer than the second peak. In both cases, the properties of the ejecta emitting the first GBM peak are different from the later ejecta, which also suggests that there might be long-timescale, a-few-second ( $\gg t_{\text{var}} \sim 1$  ms), variabilities in the outflow.

*Time-integrated spectrum.* The joint GBM–LAT spectrum of GRB 080916c can be fit by the Band function (Band et al. 1993), with peak energy around 1 MeV,  $\alpha \approx -1.0$  and  $\beta \approx -2.2$ , except for the first 4 s (Abdo et al. 2009). Because the time intervals used to construct the spectra are much longer than the MeV variability time, the resulted spectra are all time-integrated ones. Since the synchrotron self-Compton (SSC) model for GRBs, where the MeV peak is from the IC scattering by soft photon emitting electrons, predicts a bright GeV–TeV component due to the second-order up-scattering (Piran et al. 2009), no evidence for high-energy bump up to 10 GeV in observations does not favor SSC but the synchrotron model (Wang et al. 2009). In addition, the narrow  $\nu F_\nu$  spectral peak of GRB spectra favor more synchrotron emission mechanism over SSC, since SSC usually has a much broader spectral bump (Baring & Braby 2004).

In the framework of synchrotron internal shocks, our residual collision model predicts a slight spectral softening at high energies. However, due to the small detected GeV-photon number the Poisson scatter of low statistic still allows a slight softening at tens of MeV to fit the data. Furthermore, if there are large timescale,  $\gg 1$  ms, variabilities in the outflow so that the residual-emission spectral slope  $q$  is larger than  $2/3$  (see Appendix A), the high-energy spectrum is less steepened and is closer to a single power law. Finally, the slope of Equation (12) holds on average, while the later residual collisions occur between smaller number of shells, thus there might be fluctuation from this average slope.

*Highest energy emission.* The highest energy photon is detected with 13.2 GeV only 17 s after the GRB trigger. With the redshift  $z = 4.35$  this suggests GRB 080916c produces radiation up to 71 GeV in the source frame. Moreover, LAT detects 145 photons with >100 MeV, within which 14 with >1 GeV and especially only one with >10 GeV, consistent with a power-law spectrum with photon index  $\beta \approx -2$  up to  $\sim 10$  GeV scale. There might be emission extending to higher energy, say, beyond tens of GeV, from GRB 080916c following the same slope, but the detection rate is less than 1, i.e., no photon would be detected at this energy. Thus, the observations actually suggest that the high-energy spectral cutoff (or steep drop), if there exists, is more likely to be far above the energy of the only observed highest energy photon,  $\varepsilon_{\text{max}}^{\text{ob}} = 13.2 \times g$  GeV with  $g \gg 1$ . If the high-energy emission beyond 13.2 GeV is produced by internal shocks, we argue here that it may not be produced by primary internal collisions that emit MeV  $\gamma$ -rays but produced in other regions, e.g., by residual collisions.

As said above no high-energy spectral component in GRB 080916c does not favor the SSC model but synchrotron model. Now calculate the maximum synchrotron photon energy. If  $B$  is the magnetic field strength in the internal shock, the Larmor time of an electron with Lorentz factor  $\gamma$  is  $t_L' = \gamma m_e c / eB$ . The



typical particle acceleration time can be scaled by Larmor time as  $t'_a = f t'_L$  (e.g., Hillas 1984) where  $f$  is a correction factor accounting for the uncertainty of shock acceleration. It might be that  $f \gtrsim$  a few (e.g., Lemoine & Revenu 2006). In the same time the electron suffers synchrotron cooling in a typical timescale  $t'_c = 3m_e c / 4\sigma_T \gamma (B^2/8\pi)$  (we neglect the IC cooling as the IC scattering usually occurs in the deep Klein–Nishina regime for the most energetic electrons). The competition between acceleration and cooling results in a maximum Lorentz factor of accelerated electrons,  $\gamma_{\max} = (6\pi e / f \sigma_T B)^{1/2}$ . The relevant synchrotron photon frequency is a constant,<sup>7</sup>

$$\nu'_{\max} = 0.3 \gamma_{\max}^2 e B / 2\pi m_e c = 0.9 e^2 / f \sigma_T m_e c.$$

The coefficient accounts for the fact that the synchrotron power per unit frequency of an electron peaks at the frequency a fraction 0.3 of the common characteristic frequency. The maximum synchrotron photon energy  $\varepsilon_{\max}^{\text{ob}} = h\nu'_{\max} \Gamma / (1+z)$  is, therefore,

$$\varepsilon_{\max}^{\text{ob}} = 15 \Gamma_{2.5} f^{-1} (1+z)^{-1} \text{ GeV}. \quad (16)$$

Comparing the predicted maximum synchrotron energy with that implied by the observation,  $\varepsilon_{\max}^{\text{ob}} = 13.2g \text{ GeV}$ , we have a lower bound,

$$\Gamma = 1.5 \times 10^4 f_{0.5} g_{0.5}, \quad (17)$$

where the conservative values,  $f = 10^{0.5} f_{0.5}$  and  $g = 10^{0.5} g_{0.5}$ , have been taken. This bulk Lorentz factor is too large for the fireball-shock model, because it faces several problems. First, the large  $\Gamma$  leads to (primary) internal shock radius larger than the deceleration radius of GRB outflow (e.g., Lazzati et al. 1999). The deceleration radius is  $R_d \approx (E_k / 4\pi n m_p c^2 \Gamma^2)^{1/3}$ . If  $R_\gamma < R_d$ , an upper limit is obtained,

$$\Gamma < 7 \times 10^3 (E_{k,55} / n_0)^{1/8} t_{\text{var},-2}^{-3/8}, \quad (18)$$

where  $E_k = 10^{55} E_{k,55} \text{ erg}$  and  $n = 1n_0 \text{ cm}^{-3}$  are the outflow kinetic energy and medium density, respectively. Second, the large  $\Gamma$  raises problem of low-energy conversion efficiency due to slow cooling of accelerated electrons (e.g., Derishev et al. 2001). If the synchrotron cooling time of electrons with typical postshock Lorentz factor  $\gamma_m \sim m_p / m_e \sim 10^3 \gamma_3$  is required to be smaller than the dynamical time of the outflow,  $t'_c(\gamma_m) < t'_d \simeq R_\gamma / \Gamma c$ , we have

$$\Gamma < 5 \times 10^3 \gamma_3^{1/5} (L_{\text{bol}} / 10^{54} \text{ erg s}^{-1})^{1/5} t_{\text{var},-2}^{-1/5}. \quad (19)$$

In this calculation, we have assumed that the postshock magnetic field is limited by observed emission,  $B^2/8\pi \leq U_\gamma = L_{\text{bol}}/4\pi R_\gamma^2 \Gamma^2 c$ . Third, in the synchrotron internal shock models the large  $\Gamma$  leads to large collision radius  $R_\gamma \approx 2\Gamma^2 c t_{\text{var}}$ , where the magnetic field  $B$  is too small to give rise high-energy synchrotron photon energy. Using  $\epsilon_b \approx \Gamma \hbar \gamma_m^2 e B / m_e c$  and the limit  $B^2/8\pi < U_\gamma$ , the restriction to obtain synchrotron emission peaking at MeV range is

$$\Gamma < 0.4 \times 10^3 \gamma_3 (L_{\text{bol}} / 10^{54} \text{ erg s}^{-1})^{1/4} t_{\text{var},-2}^{-1/2} (\epsilon_b / 1 \text{ MeV})^{-1/2}. \quad (20)$$

<sup>7</sup> This upper bound for synchrotron energy is robust for any acceleration mechanisms involving electromagnetic processes, because the acceleration limit with  $f = 1$  is robust not only to *Fermi* shock accelerations but also to any particle accelerations through electromagnetic processes. Therefore, this bound might be valid not only to internal shock models but also to electromagnetic-dominated models.

Finally, the thermal pressure of the initial fireball is not expected to accelerate the loaded baryons to very large Lorentz factor with most energy kept as the kinetic energy of baryons. The final Lorentz factor is limited to be

$$\Gamma < 3 \times 10^3 (L_0 / 10^{54} \text{ erg s}^{-1})^{1/4} r_{0,7}^{-1/4}, \quad (21)$$

where  $L_0$  is the rate at which the central source emits energy, and  $r_0 = 10^7 r_{0,7} \text{ cm}$  is the source size (see, e.g., Waxman 2003).

The above upper limits to  $\Gamma$  imply that it may be impossible that  $\Gamma \gg 10^3$ . This appears not to match the value suggested by the highest energy band observation,  $\Gamma = 1.5 \times 10^4 f_{0.5} g_{0.5}$ , unless  $f \approx 1$  and  $g \approx 1$  are satisfied at the same time: the shock acceleration must operate at the Bohm limit; furthermore, the observed highest energy photon happens to be at the maximum synchrotron energy. Thus, the observations imply there is emission much higher than 13.2 GeV, which cannot be originated from synchrotron emission in the primary internal shocks. Actually, this high-energy emission can be produced by IC emission in residual collision shocks, as discussed in present study. So the observation of highest energy emission supports the residual collision model.

In conclusion, (1) the time delay of high-energy emission and the spectral feature of highest energy emission in GRB 080916c might provide evidences for the residual collision model; (2) its spectrum is not inconsistent with the residual emission; and (3) the time delay of  $>100 \text{ MeV}$  emission constrains the bulk Lorentz factor to be  $\Gamma \sim 300$ , a typical value usually taken. It appears to be an applicable method to determine  $\Gamma$  of GRB outflows by measuring the time delays of LAT light curves. If internal shocks also work in short GRBs, we expect similar delayed, prompt high-energy emission in short GRBs. We also caution more careful spectral analysis to find the transition between primary and residual emission.

## 5. DISCUSSION

We have considered the high-energy emission in the prompt GRB spectrum, which is dominated by the IC emission from the residual collisions. Instead of an exponential spectral cutoff, a steeper, compared to the prompt MeV emission, power-law slope  $\nu F_\nu \propto \nu^{-q}$  is expected beyond the previously thought cutoff,  $\varepsilon_0$  (Equation (5)). Here  $q$  is corresponding to the dynamical evolution of the random energy in the outflow,  $E_{\text{fluc}} \propto R^{-q}$ . In the simplified case (see Section 3.1), which is consistent with optical flash observations (LW08), we take  $q = 2/3$ , while  $0 < q \leq 2/3$  in general. The extended emission makes it complicated to detect the “cutoff energy” in the goal to constrain the GRB emission region.

Indeed, EGRET had detected prompt high-energy photons past GeV in several brightest BATSE GRBs occurring in its field of view (e.g., GRB 930131, Sommer et al. 1994; GRB 940217, Hurley et al. 1994), which suggest that the other faint GRBs may produce prompt GeV photons as well (Dingus 1995). There is also no sign of cutoff in the spectra (Dingus 1995), which, if there is, should be far exceeding  $\sim 1 \text{ GeV}$ . These EGRET results are consistent with the predicted extension of prompt emission beyond GeV. However, the cutoff is not ruled out. Given the sensitive dependence of the cutoff on  $\Gamma$  (Equation (5)), a slight variation of  $\Gamma$  in individual GRB events may lead to much higher cutoff energy,  $\gg 1 \text{ GeV}$ , explaining the prompt GeV emission in EGRET-detected GRBs. Two properties may help to

discriminate our residual-collision emission model from a very high-energy cutoff model. The first is the steepening turnover in the spectrum. For typical Lorentz factors,  $\Gamma \approx 10^2$ – $10^3$ , the expected spectral turnover is  $\sim 100$  MeV–1 TeV, well located in the windows of *Fermi* and *AGILE*. The second is the time delay of high-energy emission. One may expect systematic time delay of the high-energy photons in the residual emission model, while no delay is expected in the very high-energy cutoff model. However, the task is not easy given that for a typical event with fluence  $\sim 10^{-6}$  erg cm $^{-2}$  the observed GeV photon number is only a few. In order to have enough statistics for the spectral and temporal analyses, very bright events are needed, or one may integrate many events to obtain an average burst.

We have discussed that the recent *Fermi*-detected bright GRB 080916c might have presented a good sample. Observations do not show a simple spectral cutoff, but a spectral tail up to 70 GeV in GRB frame. More than 100 photons detected above 100 MeV obviously show a time delay of about 1 s, which can be explained by the residual IC emission and results in the determination of a typical Lorentz factor value,  $\Gamma \sim 300$ . The features of GRB 080916c support the residual-collision emission model as opposed to the very high energy cutoff model.

The high-energy emission would not extend to very high-energy. There are several effects that lead to a drop in TeV range. First, when the plasma expands to very large radius,  $R \gtrsim 2\Gamma^2 cT \sim 10^{17} \Gamma_{2.5}^2 (T/10 \text{ s}) \text{ cm}$  ( $T$  is the MeV  $\gamma$ -ray duration), there would be no overlapping between the plasma and the MeV  $\gamma$ -rays, and hence no scattering is expected. At  $R \simeq 2\Gamma^2 cT$  the cutoff energy increases to  $\varepsilon_{\text{cut}} \simeq 1(T/10 \text{ s}) t_{\text{var},-2}^{-1} \text{ TeV}$  (from Equation (11)). A lack of  $\gtrsim 1$  TeV photons would be expected for GRBs with duration  $T \lesssim 10 t_{\text{var},-2} \text{ s}$ , although it should not be an exponential cutoff. Second, in interaction with  $\sim 1$  MeV photons, the Klein–Nishina limit becomes important for electrons with  $\gamma_e \gtrsim \Gamma$ , giving rise to IC photons up to  $\varepsilon_{\text{IC}} \simeq 0.1 \Gamma_{2.5}^2 \text{ TeV}$ , beyond which the spectrum gradually turns below the low energy slope. Finally, the cosmic infrared background would absorb the  $>0.1$  TeV photons that arrive from GRBs at redshift  $z \gtrsim 0.5$ . Except for low redshift events, the observed prompt GRB emission may not extend far beyond TeV range.

It should be commented here that in the framework of the synchrotron internal shock model (Waxman 2003) we do not expect a bright high-energy component, say  $\gtrsim 1$  GeV, in the prompt emission, other than the synchrotron self-Compton model (e.g., Piran et al. 2009), therefore, the residual high-energy emission will be dominant. A detection of high-energy component in the prompt emission will be an evidence against the synchrotron model for GRBs, and vice versa. The recent *Fermi* observations of several GRBs do not support a high-energy component in GRB spectrum, since the  $>100$  MeV fluences are all less than those in MeV range (Abdo et al. 2009).

Our residual collision model is not expected to produce much longer delayed high-energy emission which is not apparently overlapped with the primary MeV  $\gamma$ -rays in times. There are more and more observations showing that GRBs produce delayed  $>100$  MeV emission even after the prompt  $\gamma$ -rays end, lasting tens or even  $\sim 10^4$  s (Hurley et al. 1994; González et al. 2003; Giuliani et al. 2008). This may require some long-lasting central engine activities or external productions (e.g., Wang et al. 2006 and references therein).

This work was partly supported by the National Natural Science Foundation of China through grants 10473010 and 10843007.

## APPENDIX A

### DYNAMICS WITH MULTI-TIMESCALE VARIABILITIES

In order to show the effect of multi-timescale variabilities on the dynamics of the outflow, we carry simulations for both cases of single- and multi-timescale variabilities for comparison. We consider a series of individual material shells  $i = 1, 2, \dots, N$ , with total shell number  $N = 3000$ , released in a duration of  $T = 3$  s, so that the interval between two nearby shells is 1 ms. The shells have equal masses but different energies, with the bulk Lorentz factor of each shell following:

$$\log \Gamma_i = 2 + \xi_i \log 9 + \phi_i, \quad (\text{A1})$$

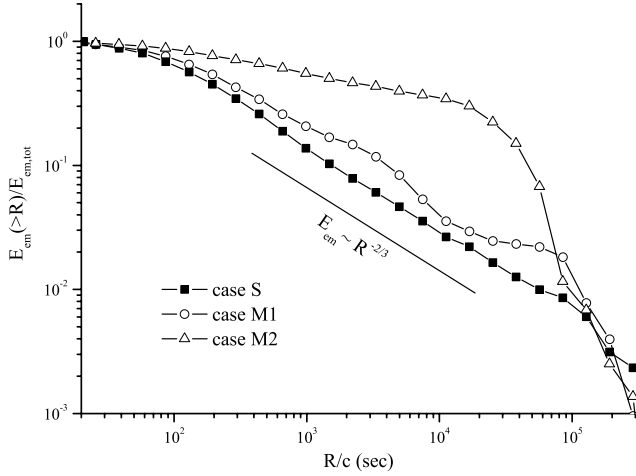
where  $\xi_i$  is a random number between zero and unity and  $\phi_i$  can be taken as the following forms,

$$\phi_i = \begin{cases} 0 & (\text{single}), \\ \sum_k \sin\left(\frac{2\pi T i}{P_k N}\right) \log A_k & (\text{multi}). \end{cases} \quad (\text{A2})$$

For single-timescale case, we should take  $\phi_i = 0$  (case S), then the outflow has only one variability timescale of  $10^{-3}$  s. The Lorentz factors are randomly and uniformly distributed in logarithmical scale between 100 and 900. For multi-timescale effect we consider three additional timescales besides the smallest ms-scale timescale:  $P_k = 10^{-2}, 10^{-1}, 1$  s for  $k = 1, 2, 3$ , with relevant  $A_k$  values being  $A_k = 2, 1.5, 1.2$  (case M1) or  $A_k = 1.2, 1.5, 2$  (case M2) for  $k = 1, 2, 3$ .  $A_k$  decreases with  $P_k$  in case M1 but increases in case M2, which means case M2 has larger fluctuations at larger timescales. Larger timescale fluctuations tend to produce strong collisions at larger radii so that larger fraction of energy is dissipated at larger radii. We further consider that in each two-shell collision, 1/3 of the generated internal energy is emitted by radiation, because only the energy of shocked electrons is assumed to be emitted rapidly and the electron equipartition parameter is taken to be  $\epsilon_e = 1/3$ . The two shells are considered to separate again after collision and share equally the remained internal energy (in the center-of-momentum frame of the two shells).

In Figure 3, we show the fraction of emission energy  $E_{\text{em}}(>R)/E_{\text{em,tot}}$ , that is emitted beyond radius  $R$ . We see that case S shows a slope close to the analytical solution  $\propto R^{-2/3}$  for single-timescale case by LW08. However, the multi-timescales lead to flatter slopes, implying more fraction of emission energy tends to be emitted at larger radii. Case M2 has even flatter slope than case M1, since case M2 has relatively larger fraction of energy that is dissipated at larger radii. The steep drop at the end in both cases M1 and M2 means no more strong collisions later on. This is because there is a largest timescale of 1 s in our simulations. If there is still variabilities with timescales larger than 1 s then the slope will continue to even larger radii and show even later drop at the end. In summary, the simulations demonstrate that multi-timescale variabilities lead to a flatter slope  $q < 2/3$ , and the  $q = 2/3$  slope is held only for the single-timescale case.





**Figure 3.** Fraction of emission energy as a function of radius in three simulations. The line marked with squares: case S, with only the 1 ms variability; circles. The other two are multi-timescale cases (M1 and M2), with three more scales of  $P_k = 10^{-2}, 10^{-1}$ , and 1 s, for  $k = 1, 2$ , and 3, respectively. The line marked with circles: case M1, with  $A_k = 2, 1.5$ , and 1.2 for  $k = 1, 2$ , and 3 (Equation (A2)); triangles: case M2, with  $A_k = 1.2, 1.5$ , and 2 for  $k = 1, 2$ , and 3 (Equation (A2)). The straight line shows the  $-2/3$  slope to guide eyes. See the relevant text for more details.

## APPENDIX B

### ANISOTROPIC IC EMISSION

Let us discuss at which energy the IC emission is emitted in the observer frame, taking into account the fact that the seed photons are beamed. Consider the extremely anisotropic case, where the photons are collinear in the rest frame of the outflow. In this frame the electrons, as argued, are reasonably assumed to be isotropically distributed. For simplicity, we consider mono-energetic photons, since the photon number rapidly decreases with energy. Thus, the IC power per unity solid angle in the rest frame is angular dependent,

$$\frac{dP'}{d\Omega'} \propto (1 - \mu')^2, \quad (\text{B1})$$

where  $\mu' = \cos \theta'$  with  $\theta'$  being the direction with respect to photon beam, and we have taken the velocity of the electron to be  $\beta'_e \approx 1$ . Hereinafter primes denote quantities in the rest frame of the outflow, while non-prime in the observer frame. Using the Lorentz transformation,  $\mu' = (\mu - \beta_\Gamma)/(1 - \beta_\Gamma\mu)$ , where  $\beta_\Gamma = (1 - 1/\Gamma^2)^{1/2}$ , we have

$$1 - \mu' = (1 + \beta_\Gamma) \frac{1 - \mu}{1 - \beta_\Gamma\mu}, \quad (\text{B2})$$

then the angular distribution of IC power in the observer frame is

$$\frac{dP}{d\Omega} = \frac{1}{\Gamma^4(1 - \beta_\Gamma\mu)^3} \frac{dP'}{d\Omega'} \propto \frac{(1 - \mu)^2}{(1 - \beta_\Gamma\mu)^5}. \quad (\text{B3})$$

This is not like the simple cone-like distribution of the isotropic-photon case. The maximum power per solid angle is emitted at an angle with

$$\mu_{\max} = \frac{5\beta_\Gamma - 2}{3\beta_\Gamma}. \quad (\text{B4})$$

The corresponding angle in the rest frame is given by

$$1 - \mu'_{\max} = \frac{2(1 + \beta_\Gamma)}{5\beta_\Gamma} \approx \frac{4}{5}. \quad (\text{B5})$$

The scattered photon energy in the rest frame is given by  $\epsilon'_{\text{IC}} \approx \gamma_e^2 \epsilon' (1 - \mu')$ , with  $\gamma_e$  being the electron Lorentz factor and  $\epsilon'$  the photon energy both in the rest frame. The photon energy (in observer frame) emitted at angle  $\mu_{\max}$ , where the IC power is maximum, is then

$$\begin{aligned} \epsilon_{\text{IC}}(\mu_{\max}) &= \frac{\epsilon'_{\text{IC}}(\mu_{\max})}{\Gamma(1 - \beta_\Gamma\mu_{\max})} = \frac{3\gamma_e^2 \epsilon'}{5\Gamma(1 - \beta_\Gamma)} \\ &\approx \frac{3}{5} \gamma_e^2 \Gamma \epsilon' (1 + \beta_\Gamma) \approx \frac{3}{5} \gamma_e^2 \epsilon, \end{aligned} \quad (\text{B6})$$

where in the last equality  $\epsilon = (1 + \beta_\Gamma)\Gamma\epsilon'$  is taken for collinear photons. This is the observer-frame photon energy around which the IC emission is mainly emitted. We see that the anisotropic correction factor  $\lambda$  in Equation (9) is only order unity, implying that taking  $\lambda \sim 1$  is a good approximation even for highly beamed seed photons.

## REFERENCES

- Abdo, A. A., et al. 2009, *Science*, **323**, 1688  
Akerlof, C., et al. 1999, *Nature*, **398**, 400  
Aoi, J., Murase, K., Takahashi, K., Ioka, K., & Nagataki, S. 2009, arXiv:0904.4878  
Band, D., et al. 1993, *ApJ*, **413**, 281  
Baring, M. G. 2000, in AIP Conf. Ser. 515, GeV–TeV Gamma Ray Astrophysics Workshop: Towards a Major Atmospheric Cherenkov Detector. VI (Melville, NY: AIP), **238**  
Baring, M. G., & Braby, M. L. 2004, *ApJ*, **613**, 460  
Baring, M. G., & Harding, A. K. 1997, *ApJ*, **491**, 663  
Beloborodov, A. M. 2000, *ApJ*, **539**, L25  
Beskin, G., Karpov, S., Bondar, S., Guarnieri, A., Bartolini, C., Greco, G., & Piccioni, A. 2009, arXiv:0905.4431  
Blake, C. H., et al. 2005, *Nature*, **435**, 181  
Blandford, R., & Eichler, D. 1987, *Phys. Rep.*, **154**, 1  
Bloom, J. S., et al. 2009, *ApJ*, **691**, 723  
D’Elia, V., et al. 2009, *ApJ*, **694**, 332  
Derishev, E. V., Kocharovskiy, V. V., & Kocharovskiy, V. V. 2001, *A&A*, **372**, 1071  
Dingus, B. L. 1995, *Ap&SS*, **231**, 187  
Giuliani, A., et al. 2008, *A&A*, **491**, L25  
González, M. M., Dingus, B. L., Kaneko, Y., Preece, R. D., Dermer, C. D., & Briggs, M. S. 2003, *Nature*, **424**, 749  
Granot, J., Cohen-Tanugi, J., & do Couto e Silva, E. 2008, *ApJ*, **677**, 92  
Greiner, J., et al. 2009a, *ApJ*, **693**, 1912  
Greiner, J., et al. 2009b, *A&A*, **498**, 89  
Gupta, N., & Zhang, B. 2007, *MNRAS*, **380**, 78  
Gupta, N., & Zhang, B. 2008, *MNRAS*, **384**, L11  
Hillas, A. M. 1984, *ARA&A*, **22**, 425  
Hurley, K., et al. 1994, *Nature*, **372**, 652  
Kobayashi, S., Piran, T., & Sari, R. 1997, *ApJ*, **490**, 92  
Krühler, T., et al. 2008, *ApJ*, **685**, 376  
Krühler, T., et al. 2009, *ApJ*, **697**, 758  
Lazzati, D., Ghisellini, G., & Celotti, A. 1999, *MNRAS*, **309**, L13  
Lemoine, M., & Revenu, B. 2006, *MNRAS*, **366**, 635  
Li, Z., & Song, L. M. 2004, *ApJ*, **608**, L17  
Li, Z., & Waxman, E. 2007, arXiv:0711.4969  
Li, Z., & Waxman, E. 2008, *ApJ*, **674**, L65 (LW08)  
Lithwick, Y., & Sari, R. 2001, *ApJ*, **555**, 540  
Lyutikov, M., & Blandford, R. 2003, arXiv:astro-ph/0312347  
Molinari, E., et al. 2007, *A&A*, **469**, L13  
Murase, K., & Ioka, K. 2008, *ApJ*, **676**, 1123  
Narayan, R., & Kumar, P. 2009, *MNRAS*, **394**, L117  
Paczynski, B., & Xu, G. 1994, *ApJ*, **427**, 708  
Pe’er, A., Ryde, F., Wijers, R. A. M. J., Mészáros, P., & Rees, M. J. 2007, *ApJ*, **664**, L1  
Piran, T., Sari, R., & Zou, Y.-C. 2009, *MNRAS*, **393**, 1107  
Racusin, J. L., et al. 2008, *Nature*, **455**, 183  
Razzaque, S., Mészáros, P., & Zhang, B. 2004, *ApJ*, **613**, 1072  
Rees, M. J., & Meszaros, P. 1994, *ApJ*, **430**, L93

- Sari, R., & Piran, T. 1997, [ApJ](#), **485**, 270
- Sommer, M., et al. 1994, [ApJ](#), **422**, L63
- Vestrand, W. T., et al. 2005, [Nature](#), **435**, 178
- Vestrand, W. T., et al. 2006, [Nature](#), **442**, 172
- Wang, X.-Y., Li, Z., Dai, Z.-G., & Meszaros, P. 2009, [ApJ](#), **698**, L98
- Wang, X.-Y., Li, Z., & Mészáros, P. 2006, [ApJ](#), **641**, L89
- Waxman, E. 2003, in *Lecture Notes in Physics 598, Supernovae and Gamma-Ray Bursters*, ed. K. Weiler, 393
- Woods, E., & Loeb, A. 1995, [ApJ](#), **453**, 583
- Wozniak, P. R., Vestrand, W. T., Panaitescu, A. D., Wren, J. A., Davis, H. R., & White, R. R. 2009, [ApJ](#), **691**, 495
- Yost, S. A., et al. 2007, [ApJ](#), **669**, 1107