Mon. Not. R. Astron. Soc. **270**, 480–498 (1994)

A model of gamma-ray bursts

Christopher Thompson
Canadian Institute for Theoretical Astrophysics, 60 St George Street, Toronto, Canada M5S 1A 1

Accepted 1994 April 19. Received 1994 April 14; in original form 1994 January 28

ABSTRACT

We present a model that reproduces the basic spectral properties of classical gamma-ray bursts with essentially no free parameters. It is an elaboration of the scenario for cosmological gamma-ray bursts outlined by Duncan & Thompson. The starting point is a Poynting-flux-dominated, relativistic, MHD wind of extremely high luminosity, $L \sim 10^{50}$ erg s^{-1}. The compactness parameter measured at the base of the wind exceeds that of the Crab pulsar, or that of a luminous AGN, by a factor of $\sim 10^{12}$. The wind emanates from a rapidly rotating neutron star, or neutron disc, in which a poloidal field $\gtrsim 10^{14}$ G has been generated by a helical dynamo. Scenarios that could produce such an object include a failed Type Ib supernova, accretion-induced collapse of a white dwarf, or perhaps a binary neutron star merger.

The wind is safely in the MHD limit as the result of neutrino-driven and centrifugally driven mass loss. Mildly relativistic Alfvén turbulence is excited in the wind by reconnection, or by hydrodynamical instabilities triggered by magnetic tension. Gamma-rays are generated via Comptonization at moderate to high scattering depth. The amplitude of the turbulence is itself limited by Compton drag, and the y-parameter of the Alfvén motions is regulated to a value near 1/4, with a weak dependence on parameters such as radius, luminosity and the amount of baryon loading. The resulting spectrum is a power law with spectral index close to $\beta = -2$ (νF_ν = constant), extending from an energy $E_{\rm break} \sim 1$ $(L_\gamma/10^{50}$ erg s$^{-1})^{1/4}$ MeV (close to the spectral peak of a thermal fireball carrying the same flux) up to an energy as high as $\sim 10^3 m_{\rm e} c^2$. This power law steepens when the amplitude of the turbulence declines, or when the turbulence is generated outside the scattering photosphere. The spectrum below energy $E_{\rm break}$ is also a power law, with index $\alpha = -1$, which is cut off from below by stimulated scattering terms.

Heavy baryon loading causes much less adiabatic softening of the spectrum than in thermal fireballs, so long as the Alfvén turbulence is generated out to the scattering photosphere. We show explicitly that the broken power law spectrum is an attractor, and that neither power law is altered by relativistic corrections to the Kompane'ets equation (except near the high-energy cut-off). The emergent gamma-ray spectrum is generated at a distance as small as $\sim 10^9$ cm from the source, without the need for any interaction with an external medium.

Key words: MHD – radiation mechanisms: non-thermal – radiative transfer – turbulence – stars: neutron – gamma-rays: bursts.

1 INTRODUCTION

The observation by BATSE that gamma-ray bursts are distributed isotropically on the sky, but with counts that fall below a Euclidean slope at low flux (Meegan et al. 1992), suggests that they are at cosmological distances[1] (Paczyński 1986; Mao & Paczyński 1992).

The sudden release of the gravitational binding energy (or rest energy) of a neutron star is easily sufficient to power a gamma-ray burst (GRB) detectable by BATSE from a redshift $z \sim 1$. The minimum energetic requirement is $\sim 10^{51}(\Delta\Omega/4\pi)$ erg if the burst is beamed into a solid angle $\Delta\Omega$. A familiar possibility is the merger of a double neutron star (or neutron star–black hole) binary by emission of gravity waves (Paczyński 1986), which will probably generate a black hole (BH) surrounded by a lower mass accretion disc (e.g. Rasio & Shapiro 1992; Davies et al. 1993). Two other scenarios are accretion-induced collapse (AIC) of a white

[1]None the less, *beamed* emission from high-velocity neutron stars in the Galactic halo can reproduce the observed spatial distribution without excessive fine-tuning of parameters (Duncan, Li & Thompson 1993).

dwarf (Duncan & Thompson 1992, hereafter DT92; Usov 1992), which creates a neutron star (NS) and usually also a disc, and the failed Type Ib supernova model proposed by Woosley (1993), in which the hydrogen-stripped core of a massive star collapses to form a black hole surrounded by a disc. Note that accretion-induced collapse may occur either via Roche lobe overflow from a non-degenerate companion, or as the consequence of a white dwarf binary coalescence.

How is the available rotational and gravitational energy converted to gamma-rays? If gamma-rays are generated near the source, then γ–γ collisions will result in a thermal pair fireball that expands to much larger radii before becoming optically thin (Paczyński 1986; Goodman 1986). This is not consistent with the highly non-thermal spectra of classical gamma-ray bursts (e.g. Band et al. 1993). There is the additional problem that a small admixture of baryons will allow the gamma-rays to be redshifted via adiabatic expansion into the X-ray range or below (e.g. Paczyński 1990). One attractive mechanism for extracting energy in electromagnetic form that could circumvent both these problems is a relativistic magnetohydrodynamic (MHD) wind (DT92). Such a wind carries both bulk kinetic energy and ordered Poynting flux, and it is possible that gamma-ray production occurs mainly at large distances from the source (DT92; Mészáros & Rees 1992; Usov 1994).

Not only is a cosmological gamma-ray burst source extremely luminous, but it is also highly *compact.*[2] In fact, the compactness parameter

$$\ell_{\rm comp} = \frac{\sigma_{\rm T} L}{m_{\rm e} c^3 R}, \tag{1}$$

which we define in terms of the spin-down luminosity L and source radius R, provides a better measure of the problem than does L alone. The value of $\ell_{\rm comp}$ exceeds those of *both* the most luminous pulsars and the most luminous AGN by some 12 orders of magnitude.

The basic question that we address in this paper is the following: how is Poynting flux dissipated around a source of such high compactness? We examine, in detail, one particular mechanism that promises to reproduce the non-thermal spectra and spectral break energies of gamma-ray bursts with essentially no free parameters. This mechanism is Compton upscattering of quasi-thermal radiation by relativistic Alfvén turbulence.

2 GENERATION OF STRONG MAGNETIC FIELDS

A rapidly rotating neutron star (or disc) releases energy via magnetic torques at the rate

$$L \sim 1 \times 10^{49} B_{15}^2 P_{-3}^{-4} R_6^6 \ {\rm erg\ s^{-1}}, \tag{2}$$

where $P = P_{-3} \times 10^{-3}$ s is the spin period, and $B = B_{15} \times 10^{15}$ G is the strength of the surface poloidal field at radius $R = R_6 \times 10^6$ cm. The field required to generate $L \gtrsim 10^{50}$ erg s^{-1} is enormous, but is easily provided by a helical dynamo operating in hot, convective nuclear matter with a millisecond spin period (DT92). A cooling neutron star must develop a negative entropy gradient when the diffusive neutrino and lepton number fluxes have settled to a quasi-steady state in the outermost stellar layers (Burrows 1987; Thompson & Duncan 1993, hereafter TD93). The resulting convective motions are extremely vigorous, with an overturn time of $\sim 1 (L_\nu / 3 \times 10^{52}\ {\rm erg\ s^{-1}})^{-1/3}$ ms. A dipole field of order 10^{15} G is weak compared to the strongest field that can be generated by differential rotation [$3 \times 10^{17} (P/1\ {\rm ms})^{-1}$ G], or by convection ($\sim 10^{16}$ G).

This dynamo model is based on direct scalings from the observed pattern of activity in convective main-sequence stars (Simon 1990). A newborn neutron star is usually a slow rotator, in the sense that the convective time in the star is small compared to a likely spin period of 10–100 ms. Such a star may support a small-scale stochastic dynamo (TD93), but not an effective helical dynamo of the α–Ω type. This is the reason why a typical pulsar does not acquire a dipole field in excess of 10^{12}–10^{13} G. It is certainly possible, however, that neutron stars sometimes do form hot *and* with initial spin periods in the millisecond range. These stars should acquire dipole fields much stronger than those of ordinary pulsars (DT92).

A similar dynamo will operate in a hot, nuclear matter disc orbiting a central neutron star or black hole. If the neutron matter rotates at the local Keplerian angular velocity, then L is approximately independent of R (neglecting general relativistic effects[3]). The required poloidal field is

$$B_{15} \simeq 0.5 L_{50}^{1/2} \left(\frac{M}{2\ {\rm M_\odot}} \right)^{-1} \tag{3}$$

for a central mass M. The surface poloidal field is almost certainly stronger at smaller radii, given the higher density of the convecting material. Indeed, the spin-down luminosity scales with poloidal flux $\Phi \sim BR^2$ as $L \propto \Phi^2/R^4$. This suggests that a neutron disc orbiting a ~ 2-M$_\odot$ Kerr black hole generates a higher luminosity than a neutron disc orbiting a neutron star. For this reason, the formation of a Kerr BH surrounded by a massive neutron disc might be favoured as a source of cosmological gamma-ray bursts over the survival of a central neutron star (which is the result of AIC). The last stable circular orbit of an extremal Kerr hole lies at a coordinate distance $R = GM/c^2 = 3(M/2{\rm M_\odot})$ km, as compared with $R = 6GM/c^2 = 18(M/2{\rm M_\odot})$ km for a Schwarzschild hole (e.g. Shapiro & Teukolsky 1983). Thus a ~ 2-M$_\odot$ Schwarzschild hole surrounded by a massive neutron disc should emit a 'spin-down' luminosity comparable to that emitted by an isolated millisecond neutron star.

A similar relativistic MHD outflow would result if angular momentum were extracted from a central Kerr hole via electromagnetic torques (Blandford & Znajek 1977). The torque may be small, however, unless the last stable orbit lies inside the ergosphere, since only in this case can the ergosphere be threaded by *open* field lines that are anchored in the (thin) disc and carry a significant Poynting flux to infinity. This requires that the angular momentum of the Kerr BH exceed 94 per cent of its maximum value (e.g. Rees 1984) in

[2]Except in models where the source is a supermassive black hole (Blandford 1993).

[3]For example, the angular velocity of a massive test particle in a circular orbit about a Kerr BH is $\Omega(R) \simeq (GM/R^3)^{1/2} [1 + (J/J_{\rm max}) (GM/Rc^2)^{3/2}]^{-1}$, as measured by a clock positioned far from the hole (Shapiro & Teukolsky 1983). Here, J is the angular momentum of the hole, and $J_{\rm max} = GM^2/c$.

order to generate a spin-down luminosity comparable to (2). It is not clear that a merged neutron star binary will retain this much angular momentum after the new BH–disc system has settled down to axisymmetry, and the extraction of angular momentum by gravity waves has stopped. For example, the simulation of Davies et al. (1993) shows that the central merged star of mass $\sim 2.6\ M_\odot$ has an angular momentum of 3×10^{49} erg s, which amounts to only ~ 50 per cent of the critical angular momentum of a Kerr hole of that mass. The surrounding disc carries only ~ 10 per cent of the mass, and most of its angular momentum will presumably be carried to infinity by viscous or magnetic torques. More angular momentum can be generated in the failed Type Ib supernova scenario; see Section 2.2 If the central black hole does acquire $J > 0.94\, J_{max}$ then there is the interesting possibility that sudden accretion of a relatively small mass from an extended disc could trigger a repeat electromagnetic outburst long (perhaps hours or days) after the binary merger event.

2.1 Is convection necessary?

A neutron disc is likely to be convective if the accretion luminosity is higher than $\sim 10^{50}$–10^{51} erg s^{-1}. The disc is then hot enough to be optically thick to neutrinos, and convective instability is a direct consequence of the hot nuclear equation of state (TD93). Even if the accretion luminosity is lower, a massive disc (such as could form in the failed Type Ib supernova model: Woosley 1993) would undergo a brief convective period as a result of secular cooling. By contrast, the disc formed in a NS–NS merger carries only ~ 10 per cent of the total mass and is formed quite cool, $T \sim 2$–4 MeV, according to the smoothed particle hydrodynamics simulations of Davies et al. (1993). This slightly exceeds the internal temperature at which a cooling neutron star becomes optically thin to neutrinos, and convection ceases. Whether or not such a disc is convective therefore depends on complicated details of the merger physics that are still poorly understood (for a recent discussion, see Rasio & Shapiro 1994). If a period of convection is a necessary step in the formation of a strong, large-scale poloidal field, the survival of a massive, rapidly rotating neutron star (for at least 10–100 s) might be favoured as the end-point of a binary NS merger over the prompt formation of a BH (DT92). Whether a neutron star survives clearly depends on the hardness of the nuclear equation of state and the mass of the merging stars.

Is convection needed to generate a strong, large-scale poloidal field in a rotationally supported disc? In the absence of convection, the magnetic shearing instability (Balbus & Hawley 1991) will amplify the disc's internal magnetic field. Recall that this instability is powered by a release of shear kinetic energy (which is immediately replenished as the disc spreads in the central gravitational potential), whereas convection in a new-born neutron star is driven by *secular* neutrino cooling. It is not obvious, however, that the magnetic shearing instability can generate a mean poloidal field as strong as (3), since to first order it does not amplify the total magnetic flux threading the disc. The non-linear evolution of the instability depends sensitively on details of magnetic reconnection (Goodman & Xu 1993). A plausible disc dynamo mechanism that assumes vertical convective stability was outlined by Tout & Pringle (1992). This mechanism, however, assumes that the magnetic field is strong enough (1) to overcome the vertical stable stratification and rise buoyantly out of the disc, and (2) to generate a Balbus–Hawley mode with a vertical wavelength comparable to the disc thickness.[4] In a Keplerian disc with a mass of order 1 $M_\odot$ orbiting a neutron star or ~ 2-$M_\odot$ black hole, this requires a field stronger than $B_{buoy} \sim 10^{17}$ G. Otherwise, the buoyant motions of magnetic flux ropes through the nuclear matter are effectively suppressed (cf. Goldreich & Reisenegger 1992).

This leads to the key question: can the large-scale poloidal field be amplified to the value (3), without convection, starting from a *very weak* seed field? Recent empirical work on pulsar field decay (Wijers et al. 1993) suggests that the strong dipole fields of isolated pulsars decay by only a moderate factor, if at all, on time-scales as long as $\sim 10^7$–10^8 yr. The field anchored in the deep crust is limited by turbulent Hall drift to a strength $\sim 5 \times 10^{11}\ (t_{merge}/10^9\ \mathrm{yr})^{-1}$ G (Goldreich & Reisenegger 1992). Thus binary neutron stars that merge on a time-scale $t_{merge} \sim 10^9$–10^{10} yr plausibly provide a seed field of $B_0 \sim 10^{11}$–10^{12} G.[5] This is still 3–4 orders of magitude weaker than the poloidal field (3) and 5–6 orders of magnitude weaker than B_{buoy}. The azimuthal field would be amplified by the disc shear alone to a strength B_{buoy} in $(B_{buoy}/2\pi B_0)$ rotation periods, or equivalently in $16 P_{-3} (B_{buoy}/10^{17}\ \mathrm{G})(B_0/10^{12}\ \mathrm{G})^{-1}$ s. The resulting azimuthal field would reverse sign on a very short length-scale, $\lambda/R \sim 2\pi B_0/B_{buoy}$. This is also comparable to the scale of the fastest growing Balbus–Hawley (1991) mode. It has been suggested that magnetic reconnection can smooth reversals in the field on very small scales, pushing the dominant growing mode to much larger scales (cf. Goodman & Xu 1993). The dissipation rate due to reconnection in a highly conducting, incompressible fluid is, however, severely limited by the magnetic tension, which suppresses turbulent fluid motions on small scales (Cattaneo & Vainshtein 1991; TD93).

A model of gamma-ray bursts based on magnetic disc flares has been sketched by Narayan, Paczyński & Piran (1992). This model can accommodate a much smaller scale surface field than the magnetic torque mechanism described above. (Note that, even in this model, it is possible that the required strong magnetic field is built up in the neutron disc by a helical dynamo during a transient period of convection.) What, then, is the energy flux from the surface of the disc in the form of reconnected magnetic field lines and entrained plasma? How does this energy flux compare with the rotational energy loss (2)?

Since a calculation from first principles is not feasible, we must resort to phenomenological scalings. Magnetic activity

[4]The buoyant motions of magnetic flux tubes convert azimuthal and radial fields to vertical field, and the magnetic shearing instability converts vertical field to radial field (Tout & Pringle 1992). The cycle is closed in the usual manner by the disc shear, which converts radial field to azimuthal field.

[5]The field strength could be reduced further by ohmic decay, depending on the concentration of impurities in the crust, as well as the amount of penetration of the field into the core (e.g. Sang & Chanmugam 1987). The field anchored in the core could be stronger, depending on details of magnetic flux transport that are not fully understood (e.g. Goldreich & Reisenegger 1992).

in rapidly rotating stars with convective envelopes is directly tied to X-ray emission and flare activity (e.g. Haisch, Strong & Rodonò 1991). The time-averaged kinetic energy luminosity in the solar wind is $\sim 10^{-8}$ $L_{\odot}$, of which an appreciable fraction is in the form of discrete ejections of mass and magnetic field energy ('coronal mass ejections'). We denote by L_{fl} the time-averaged luminosity of the mass and magnetic field ejected from a system. This quantity is of greater interest than the quiescent X-ray luminosity L_X, which has little meaning when the source has a high compactness. For example, the photons advected by magnetic flux ropes above the neutrinosphere of a cooling neutron star are effectively trapped in the ropes, due to the high scattering optical depth.

Very rapidly rotating, late-type stars flare much more frequently than the Sun. The time-averaged X-ray flare luminosity of short-period binary G–F dwarfs (RS CVns: Hall 1989) is comparable[6] to the quiescent X-ray luminosity (e.g. Stern et al. 1991). Although one does not know how often RS CVn flares are preceded by mass expulsion, this would suggest that the time-averaged luminosity L_{fl} of the mass and magnetic field ejected from these systems is comparable to the quiescent X-ray luminosity L_X. The L_X values measured for RS CVn systems still do not exceed $\sim 10^{-3}$ of the bolometric luminosity (Pallavicini et al. 1981). The X-ray emission from a rotationally supported disc, however, can be a much larger fraction of the accretion luminosity of the disc. For example, the rapidly variable X-ray emission of Seyfert 1 galaxies (which is almost certainly generated in the inner regions of an accretion disc) amounts to ~ 10 per cent of the bolometric luminosity (Mushotzky, Done & Pounds 1993; see also Section 5.1). If $L_{fl} \sim L_X$ in these discs, then it would be reasonable to suppose that a neutron disc accreting on to a ~ 2-$M_{\odot}$ black hole would also generate $L_{fl}/L_{\nu} \sim 10$ per cent.

Whatever the value of L_{fl}, a disc orbiting a near-extremal Kerr black hole does emit a substantial 'spin-down' luminosity. This luminosity is, to within an order of magnitude, the magnetic field energy outside the disc divided by the light crossing time of the hole: $L \sim (B^2/8\pi) \times 2\pi (GM/c^2)^2 c$. Reconnection would need to be very efficient in order to compete with this energy loss mechanism.

2.2 Angular momentum requirements

The failed Type Ib supernova scenario (Woosley 1993) can provide a much more massive neutron disc than a binary NS merger. In this scenario, the pre-collapse H-stripped core is necessarily a fast rotator. The simplest way to generate the required angular momentum is for the core to reside in a tight binary (in which the companion star is probably a neutron star) and be in corotation with the binary. The requirement that the total core angular momentum exceed the maximum angular momentum of a Kerr BH of the same mass places interesting limits on the binary period. We make the reasonable assumption that the binary is circular, and model the core as an $n=3$ polytrope.[7] Then the binary period must be *smaller* than

$$P_{orb} = 1.5 \left(\frac{M_{core}}{2\ M_{\odot}}\right)^{-1} \left(\frac{R_{core}}{10^{10}\ cm}\right)^{2}\ h. \qquad (4)$$

This orbit is tight enough that the core may in fact have been stripped of its helium in a common envelope to form a CO core–NS binary. The formation rate of such tight binaries is presumably quite low, somewhat less than the formation rate of the He core–NS binaries that were the precursors of the double NS systems 1913+16 and 1534+12.

Alternatively, it may sometimes happen that the core of a very massive star retains the required angular momentum as its outer hydrogen layers are blown off in a stellar wind. Indeed, Brown & Bethe (1994) have noted that the cores of stars more massive than $\sim 20\ M_{\odot}$ will undergo a (delayed) collapse to form a black hole if the nuclear equation of state is soft. Prompt formation of a BH introduces a mechanism for failure of a core collapse supernova, if the success of the shock depends on delayed neutrino heating (as suggested by Wilson 1985). The corresponding rate of Kerr hole formation depends, of course, on the physics of angular momentum transport inside the progenitor star. The required spin could also be generated when the core merges with a binary companion (or the core of a binary companion) during a common-envelope phase.

Finally, we should emphasize that the end product of AIC is almost always a star that has too much angular momentum to collapse directly to nuclear density – a 'fizzler' (Narayan & Popham 1989). Such a star may become de-leptonized at a density below the neutrino trapping density. In this case, it is not clear that the resulting neutron star will be heated sufficiently to develop a convective instability, even though it is a rapid rotator. Some AIC events, however, will produce stars that are *both* hot (collapse directly to nuclear density) and rapid rotators, and are the sites of an effective $\alpha-\Omega$ dynamo (DT92). Rotation will substantially reduce the strength of the bounce shock if the collapse is halted by centrifugal forces *inside* the mass shell where the shock would first appear in the absence of rotation. In order for this to happen, the progenitor white dwarf must have a dipole field *weaker* than

$$B_{wd} \sim 2 \times 10^{6} \left(\frac{\dot{M}}{10^{-6}\ M_{\odot}\ yr^{-1}}\right)^{1/2}\ G \qquad (5)$$

when its mass reaches the Chandrasekhar mass M_{Ch} (Thompson & Duncan 1994). We assume that the magnetized dwarf spins at the equilibrium period corresponding to the accretion rate $\dot{M}$ (e.g. Bhattacharya & van den Heuvel 1991).

Note also that, if the field were *not* amplified in the newborn neutron star, a dipole field of $\sim 10^{15}$ G would require a dipole in the pre-collapse white dwarf of $\sim 4 \times 10^{10}$ G. The initial neutron star spin period would lie well in excess of 1 ms, even if the progenitor dwarf were accreting at the Eddington rate of $\dot{M} \sim 2 \times 10^{-6}\ M_{\odot}\ yr^{-1}$. This contradicts the suggestion made by Usov (1992) that such a neutron star could be a millisecond rotator.

[6]By comparison, the quiescent X-ray luminosity $L_X \sim 10^{-6}$ $L_{\odot}$ emitted by hot, magnetically confined plasma in the solar corona greatly exceeds L_{fl} (e.g. Simon, Herbig & Boesgaard 1984).

[7]The relative incompressibility of nuclear matter above nuclear density implies that the density profile of a neutron star is much less centrally concentrated than that of a white dwarf. A neutron star of mass M and radius R has a moment of inertia $I \simeq 0.35\ MR^2$ (Arnett & Bowers 1977), as compared to $I \simeq 0.075\ MR^2$ for an $n=3$ polytrope.

3 MASS LOSS AND SCATTERING DEPTH

3.1 Neutrino-driven mass loss

To begin, consider a steady, spherical wind composed of Poynting flux and a small scattering contaminant of electrons and ions. Electron–positron pairs should dominate the opacity very close to the source, but that will not be important for our discussion. The luminosity of the wind is assumed to be very high, $L = L_{50} \times 10^{50}$ erg s^{-1}, and since the source is very rapidly rotating (at period $P_{-3} \sim 0.1$–1) we may define the effective source radius to be the light-cylinder radius $r_{lc} = cP/2\pi = 4.8 \times 10^6 \ P_{-3}$ cm. A lower bound to the photon-pair flux may be estimated from the reaction $\nu + \bar{\nu} \to e^+ + e^-$, which yields $L_\gamma(R_\nu) \sim 10^{-3} L_\nu$ (Janka 1991). This corresponds to $L_\gamma(R_\nu) \sim 3 \times 10^{49}$ erg s^{-1} averaged over the first 3 s of the cooling history of a typical neutron star. We emphasize that $L_\gamma(R_\nu)$ is the photon-pair luminosity just outside the neutrinosphere. The photon luminosity further out in the wind may be a much larger fraction of the total spin-down luminosity. For example, turbulence in the wind is damped effectively by Compton drag (Section 5).

The high neutrino flux ablates baryonic material from the stellar surface. A rest mass flux $\dot{M}_b$ limits the bulk Lorentz factor of the wind to a value

$$\gamma_b = \frac{L}{\dot{M}_b c^2} = 5.6\, L_{50} \left(\frac{\dot{M}_b}{10^{-5}\ M_\odot\ s^{-1}} \right)^{-1}. \tag{6}$$

We note that formula (2) for the spin-down luminosity applies only in so far as $\dot{M}_b < L/c^2 (\gamma_b \gtrsim 1)$. The most detailed calculation of such a neutrino-driven wind has been made by Witti, Janka & Takahashi (1994). They find

$$\dot{M}_b \simeq 2 \times 10^{-3} \left(\frac{L_\nu}{3 \times 10^{52}\ \mathrm{erg\ s^{-1}}} \right)^2 M_\odot\ s^{-1}, \tag{7}$$

using a simple parametric cooling law for neutrino luminosities in the range $3 \times 10^{51} \lesssim L_\nu \lesssim 3 \times 10^{52}$ erg s^{-1}. Most of the ablation is due to neutrino absorption on nucleons (Duncan, Shapiro & Wasserman 1986). Indeed, at $L_\nu \sim 3 \times 10^{52}$ erg s^{-1}, the ablation rate (7) is ~ 20 times higher than what can be driven, on energetic grounds, by $\nu - \bar{\nu}$ annihilation alone. [To see this, combine the pair-photon luminosity calculated by Janka (1991) with the estimate $\dot{M}_b \sim L_\gamma(R_\nu) \times (GM_*/R_*)^{-1} \simeq 5 L_\gamma(R_\nu)/c^2$]. Mass-loss rates as large as (7) are consistent with the amount of pressure stratification observed in the high-entropy bubble that forms outside the neutrinosphere in (both one- and two-dimensional) supernova simulations (e.g. Miller, Wilson & Mayle 1993).

Extrapolation of (7) to the lower luminosities characteristic of the last stages of optically thick neutrino cooling ($L_\nu \sim 10^{50}$–10^{51} erg s^{-1} at ~ 30 s after bounce for a non-rotating neutron star) suggests a much lower mass-loss rate, $\dot{M}_b \simeq 2 \times 10^{-7} \ (L_\nu / 3 \times 10^{50}$ erg s$^{-1})^2$ $M_\odot$ s^{-1}, and a correspondingly high limiting Lorentz factor, $\gamma_b \sim 300 \ (L_\nu / 3 \times 10^{50}$ erg s$^{-1})^{-2}$. If we assume that the external poloidal field strength B_p is limited by the vigour of the convective motions, then the spin-down luminosity scales with neutrino flux (at fixed rotation period) as $L \propto B_p^2 \propto V_{con}^2 \propto L_\nu^{2/3}$. (Here V_{con} is the convective velocity.) The limiting bulk Lorentz factor of the wind *increases* as the neutrino luminosity drops off: using the scaling (7), we have

$$\gamma_b = \frac{L}{\dot{M}_b c^2} \propto L_\nu^{-4/3}. \tag{8}$$

This suggests that, when the effects of baryon contamination are taken into account, the gamma-ray luminosity is *bounded above*: there is a maximum neutrino flux (and therefore a maximum value of L) above which the gamma-rays are smothered and adiabatically degraded before they can escape the wind. A specific example is provided by the reconnection model outlined in Section 4.

The neutrino luminosity emitted when disc material of mass ΔM_{disc} accretes via viscous (or magnetic) torques on a time-scale $\Delta t \sim 1$–100 s is

$$L_\nu \sim 1.7 \times 10^{52} \left(\frac{M_{disc}}{M_\odot} \right) \left(\frac{\Delta t}{10\ s} \right)^{-1} \mathrm{erg\ s^{-1}} \tag{9}$$

for a canonical radiation efficiency[8] of 0.1. If we demand that the neutrino luminosity be less than 10^{51} erg s^{-1} in order to suppress the baryon ablation rate, the total accreted mass

$$\Delta M_{disc} \lesssim 0.06 \left(\frac{\Delta t}{10\ s} \right) \tag{10}$$

is very small (unless the burst is very long). We emphasize that ΔM_{disc} represents the change in the disc mass; the total disc mass could be much larger, as could the initial disc mass. None the less, this suggests that there is a considerable delay between the initial catastrophic energy release and the onset of the gamma-ray burst in a scenario where the disc is initially massive (such as the failed Type Ib supernova model: Woosley 1993).

The gamma-ray luminosity emitted by a magnetized neutron disc is self-limiting, in the manner described above for a cooling neutron star. For example, if 10 per cent of the accretion luminosity of the disc is converted into Poynting flux, and if L_ν is required to be lower than $\sim 10^{51}$ erg s^{-1} in order to keep γ_b above the critical value (24), then L_γ is limited to $\sim 10^{50}$ erg s^{-1}. Somewhat higher gamma-ray luminosities may be possible if the wind is emitted by a hydrostatically supported star rather than by a rotationally supported disc: the rotational energy of the star can be tapped without converting gravitational binding energy to heat, as must happen in the disc.

3.2 Magnetic trapping and centrifugal effects

The surface poloidal field B_p that enters into the spin-down formula (2) is obtained by assuming that the field between the surface of the star (or disc) and the light cylinder is purely dipolar. A stronger, high-multipole field is probably present in a hot, convective neutron star (TD93). This multipolar field can be visualized as a large number of densely packed flux loops that leave and re-enter the star at closely separated points. Plasma inside these flux loops is effectively trapped near the stellar surface (TD93). The ram pressure of the expanding baryons can be contained by a field as weak as

$$B \sim (\dot{M} V_{esc})^{1/2} / R_* \sim 3 \times 10^{13} (\dot{M}/10^{-5}\ M_\odot\ s^{-1})^{1/2} \times (R_*/10\ \mathrm{km})^{-5/4}\ \mathrm{G},$$

[8]The efficiency of conversion of rest mass to radiation could be as high as ~ 0.4 for a Kerr black hole: e.g. Shapiro & Teukolsky (1983).

assuming that the baryons move at the escape velocity V_{esc}. This field is much weaker than the rms field $B_{rms} \sim 10^{16}$ G allowed by convection, or the field $B_{rms} \sim 3 \times 10^{17}$ G allowed by differential rotation, in an isolated neutron star (TD93). (The limiting small-scale field will be even stronger in a high-mass neutron star remnant, or in neutron matter orbiting close to a $\sim 2\text{-M}_\odot$ Kerr black hole.) The fraction of the field lines that open out to infinity is B_p/B_{rms}, and the net neutrino-driven mass-loss rate is reduced from the spherical estimate by this fraction. For example, a luminosity $L \sim 10^{50}$ erg s^{-1} emitted by a neutron disc orbiting a central mass of 2 $M_\odot$ requires $B_p \sim 5 \times 10^{14}$ G, which in turn implies that the *steady* mass-loss rate is reduced by a factor $\sim 20(B_{rms}/10^{16}\text{ G})$. The mass loss would be further suppressed if the wind were collimated into a jet: in this case, the lower spin-down luminosity could be provided by a weaker poloidal field. The occasional reconnection of small-scale flux tubes will inject dense baryon clouds into the wind, but these are disconnected from most of the expanding field lines, and need not retard them appreciably.

Mass loss from a disc is driven by centrifugal acceleration along open field lines (Blandford & Payne 1982) as well as by neutrino heating. Efficient acceleration from a thin (Keplerian) disc requires that the field lines be inclined at less than 60° with respect to the equator. This suggests that centrifugally driven mass loss will be heaviest in the outer parts of the disc, and that a detectable gamma-ray signal may be emitted only within a certain solid angle centred on the rotation axis. In general, centrifugal acceleration becomes easier as a disc spreads and thins out, and its angular velocity comes closer to the Keplerian value.

We conclude that a baryon loading as low as $\gamma_b \sim 10^2$ requires a very low neutrino luminosity, but one that is reached during the last stages of diffusive neutrino cooling. The trapping of baryons by a strong surface magnetic field can increase γ_b by an order of magnitude or more. The neutrino luminosity is also suppressed in scenarios where neutronization occurs at a low density (as is expected in most AIC events: Thompson & Duncan 1994), or where it was completed long before the cataclysmic event that triggers the gamma-ray burst (as in a binary neutron star merger). The absence of a convective instability may preclude the strong, large-scale poloidal field needed to provide the spin-down torque.

3.3 Scattering depth

The very high baryon loss rate from a cooling neutron star/disc has an important consequence: the electromotive force generated by unipolar induction is effectively quenched. A gamma-ray luminosity L_γ at the source implies an extremely high charged particle flux,

$$\begin{aligned}\dot{N}_{e^-} + \dot{N}_p &= (1 + Y_e)\,\dot{M}_b/m_n \\ &\simeq L_\gamma(R_\nu)/m_n c^2 \\ &= 3 \times 10^{53} [L_\gamma(R_\nu)/10^{50}\ \text{erg s}^{-1}]\ \text{s}^{-1}. \end{aligned} \tag{11}$$

This flux is *much* higher than that required to maintain the spin-down torque. Writing the spin-down luminosity in terms of a current I, we have $I/e \simeq (Lc)^{1/2}/e \simeq 3 \times 10^{39} L_{50}^{1/2}$ s^{-1}. In the language of pulsars, the magnetospheric charge density greatly exceeds the corotation charge density (Goldreich & Julian 1969). One does *not* expect efficient gamma-ray production via unipolar induction, as suggested by Usov (1992). Conversion of Poynting flux to gamma-radiation is more likely to occur at large distances from the source, well outside its light cylinder (DT92).

The scattering optical depth in the wind rest frame is very high at $r \sim r_{lc}$ (cf. Paczyński 1990; Shemi & Piran 1990). Contamination by baryons has traditionally been viewed as a disease to be avoided in cosmological GRB models, but in fact a much heavier baryon loading can be tolerated if most of the gamma-rays are generated near the Thomson photosphere (Section 5.5). The contribution to the Thomson optical depth from the baryon/electron contaminant alone is

$$\frac{d\tau_{es}}{d\ln r} = 2.4 \times 10^{10} \gamma^{-2} \left(\frac{r}{r_{lc}}\right)^{-1} \left(\frac{Y_e}{\gamma_b}\right) L_{50} P_{-3}^{-1}. \tag{12}$$

Here Y_e has its usual definition of electron number per baryon, and we approximate the wind as spherical. Let us make the further approximation that the bulk Lorentz factor increases linearly with radius as $\gamma \simeq r/r_{lc}$ up to the limiting value $\gamma \simeq \gamma_b$. The wind becomes optically thin to electron scattering *before* the baryon inertia begins to limit γ if

$$\gamma_b > \gamma_b^* = 3.9 \times 10^2 (Y_e L_{50}/P_{-3})^{1/4}. \tag{13}$$

The quantity γ_b^* represents the largest bulk Lorentz factor achievable inside the Thomson photosphere of a freely expanding wind. Optical depth unity is reached at a radius

$$r_\tau = \gamma_b^* \left(\frac{\gamma_b}{\gamma_b^*}\right)^{-1/3} r_{lc} = 1.9 \times 10^9 (Y_e L_{50})^{1/4} P_{-3}^{3/4} \left(\frac{\gamma_b}{\gamma_b^*}\right)^{-1/3}\ \text{cm} \tag{14}$$

when $\gamma_b > \gamma_b^*$, and

$$r_\tau = \gamma_b^* \left(\frac{\gamma_b}{\gamma_b^*}\right)^{-3} r_{lc} \tag{15}$$

when $\gamma_b < \gamma_b^*$. Given our previous estimate of the rest-mass flux at the source, a relatively low baryon loading $\gamma_b \sim \gamma_b^*$ would require either a low neutrino flux or a high Poynting flux. For example, $L \sim 10^{52}$ erg s^{-1} could easily be achieved by a neutron torus with $B \sim 5 \times 10^{15}$ G circulating with a minimum period $P \sim 10^{-4}$ s around a $2\text{-M}_\odot$ extremal Kerr BH. There is the danger, however, that the gamma-ray signal from such a luminous wind would be too high and, indeed, baryon loadings heavier than $\gamma_b \sim \gamma_b^*$ can be tolerated in the radiative model that we discuss in the next section.

In models that reproduce the observed hard, non-thermal gamma-ray spectrum, one expects that non-thermal electron pairs will contribute to the scattering opacity at the radius where the gamma-rays are generated. The equilibrium pair density is obtained by balancing the annihilation rate of (cold) pairs, $\dot{n}_{e^+} = -(3/8)\,\sigma_T c n_{e^+}^2$, with the production rate.

A minimal pair creation rate is obtained by taking the observed spectrum with high-energy index $\beta \simeq -2$ at $E_\gamma > E_{soft}$, and assuming that this index extends up to a rest-frame energy $E_\gamma \sim m_e c^2$. Using the pair production opacity calculated by Svensson (1987), one finds that $2n_e^+ \simeq n_\gamma$, where $(4/3)\gamma^2 m_e c^2 n_\gamma \simeq L_\gamma/4\pi r^2 c\ \ln(m_e c^2/E_{soft})$ and n_γ is the photon density at energy $\sim m_e c^2$. Converting the pair density to an optical depth, one finds

$$\frac{d\tau_{es}(e^\pm)}{d\ln r} = 5 \times 10^{12} \gamma^{-3} \left(\frac{r}{r_{lc}}\right)^{-1} (\varepsilon_\gamma L_{50}) P_{-3}^{-1}. \tag{16}$$

Here ε_γ is the fraction of the Poynting flux converted *locally* to gamma-rays, per logarithmic radius. We see that this scattering depth exceeds the depth (12) due to the electron–baryon component by a factor $\sim 200(\varepsilon_\gamma/Y_e)$. As a result, γ_b^* increases by a factor $3.8(\varepsilon_\gamma/Y_e)^{1/4}$.

4 RECONNECTION IN AN ULTRALUMINOUS MHD WIND

The Poynting flux carried by the wind may be viewed, in the ideal MHD limit, as a magnetic field advected by the charged particle component. Under the assumption of axisymmetry, this field is predominantly toroidal at $r \gg r_{lc}$ (e.g. Weber & Davis 1967; Goldreich & Julian 1970). The field strength in the wind rest frame is

$$B_r = 1.7\times 10^{13}\gamma^{-1}(r/r_{lc})^{-1}L_{50}^{1/2}P_{-3}^{-1}\,\mathrm{G} \tag{17}$$

under the simplifying assumption that the wind is spherically symmetric.[9]

Existing solutions of the MHD wind equations usually assume that the poloidal field at the base of the wind has a constant sign (at least in each hemisphere), and neglect complications associated with current sheets. We are interested here in more complicated geometries that will allow one or multiple current sheets. One simple example is a disc whose magnetic field is generated by an internal dynamo, the sign of which is random and changes on a radial length-scale of order the disc height, and a time-scale that is some multiple of the convective (or rotation) period. The sign of B_ϕ varies on a length-scale $\sim cP/2 = \pi r_{lc}$ out to large distances $r \gg r_{lc}$. Discontinuities in the poloidal field would be easiest to sustain in a disc orbiting a near-extremal Kerr black hole: the period of the last stable orbit is not much longer than the light travel time across the hole. The magnetic shearing instability (Balbus & Hawley 1991) will cause an even faster variation in the sign of the field.

Dissipation of magnetic energy at current sheets in relativistic winds has been considered by Coroniti (1990) in the context of pulsar winds, and by Romanova & Lovelace (1992) in the context of extragalactic jets. We now discuss how this physics is modified when the compactness parameter (1) is increased by more than 12 orders of magnitude from that of the Crab pulsar wind or the outflow from a luminous AGN.

4.1 Relativistic bulk motions

A reconnection event induces bulk mass motions in the wind rest frame. Consider the basic Petschek (1964) solution to the reconnection problem: two domains containing magnetic field of opposite sign are separated by a thin ohmic layer and bounded by MHD shocks. Plasma in the two domains streams towards the ohmic layer at a velocity that is a moderate fraction of the Alfvén velocity. From a more global perspective, the reconnection process involves a change in the connectedness of the field lines, and is driven by the release of magnetic tension and curvature energy. Reconnection also occurs when the particle rest energy density is small compared to the magnetic pressure, and the bulk motions are relativistic (Semenov & Bernikov 1991; Field & Rogers 1993). In this relativistic limit, the reconnecting field lines accelerate the plasma to a Lorentz factor

$$\gamma_A \simeq (\gamma_b/\gamma)^{1/2} \tag{18}$$

in the wind rest frame (so long as the expansion of the wind is not yet limited by the baryon inertia, $\gamma < \gamma_b$). This is the Lorentz factor at which a kink in the magnetic field propagates along the field lines.

These bulk motions are rapidly decelerated by Compton drag to a Lorenz factor

$$\gamma_A \sim \left(\varepsilon_\gamma \frac{d\tau_{es}}{d\ln r}\right)^{-1/2} \tag{19}$$

at low scattering depth, $d\tau_{es}/d\ln r \ll 1$. Here, ε_γ is the ratio of the photon energy density to the ambient magnetic energy density. A reconnection event generates quasi-linear Alfvén turbulence at higher optical depths, where $d\tau_{es}/d\ln r \gtrsim \varepsilon_\gamma^{-1}$. Such turbulence could also be excited by hydrodynamical instabilities (for example, if the MHD wind were collimated into a jet: DT92).

4.2 Opacity-limited reconnection

There are two physical novelties here. First, dissipation of magnetic energy at a neutral sheet is strongly inhibited near the light-cylinder by the high scattering opacity, which prevents the dissipated energy from diffusing away from the neutral sheet. The build-up of thermal pressure at the neutral sheet in turn prevents the advection of more flux to the sheet. Secondly, the stabilizing effects of ion pressure are absent (except very far from the source) due to rapid electron–ion equilibrium and synchroton cooling. This suggests[10] that in regions of low optical depth the volume of annihilated flux grows linearly with time, not quadratically as with diffusion, because the fluid in which the flux has dissipated cools and collapses against the external pressure. In regions of high optical depth, the annihilated volume grows diffusively, but the diffusivity is essentially $\frac{1}{3}\ell c$, where ℓ is the photon mean free path.

This second effect may seem counterintuitive at first sight, since the plasma conductivity is so high. If the fireball fluid were adiabatic and/or incompressible, then the magnetic dissipation rate would indeed be tiny, due to the suppression of small-scale fluid motions by the strong magnetic tension (Cattaneo & Vainshtein 1991; TD93). When the ther-

[9] We will assume spherical symmetry unless otherwise indicated, although considerable violation of this assumption may occur in practice. For example, cross-field stresses can induce considerable collimation into a jet directed along the rotation axis, and self-similar wind solutions suggest that the cylindrical radius of the jet is limited to $\varpi/r_{lc} \sim \gamma_b$ (Li, Chiueh & Begelman 1992). The scalings of physical quantities with radius and the numerical values of the various photospheres will change as a result of such collimation (the most important difference possibly being the scaling of bulk Lorentz factor with radius), but we do not expect our main conclusions to be altered qualitatively. The first-order effect of collimation is to reduce the energetic requirement for any single burst source, and to increase the number of sources required to match the observed burst rate.

[10] The cancellation of magnetic flux at a neutral sheet is not understood from first principles, but observations of solar flares suggest that the non-linear evolution of the tearing mode instability generates strong plasma turbulence and results in much more rapid flux cancellation than would be expected from a linear analysis (e.g. Priest & Forbes 1992).

mal energy generated at a current sheet is rapidly radiated away, however, reconnection is greatly speeded up. The baryon contaminant is compressed by as much as a factor $\sim\gamma_b(m_n c^2/kT)\gg 1$ before the particle pressure becomes comparable to the ambient magnetic pressure. The net result is that reconnection in a relativistic outflow of this luminosity is much *more* efficient than, say, in the solar wind.

We emphasize that the suppression of magnetic dissipation at high scattering optical depth is ultimately a result of the *topology* of the neutral sheet combined with the *boundary conditions* on the field. Point reconnection, of the sort first discussed by Petschek (1964), is certainly not suppressed when the scattering depth through the fluid masses surrounding the neutral point is high. Point reconnection requires essentially no cancellation of magnetic flux, merely a change in the connectedness of the field lines. As a result, one expects that neighbouring flux loops at the surface of a hot, convective neutron star will reconnect with each other quite easily (TD93), even though the field is extremely strong, $B\sim 10^{16}$ G. Efficient reconnection at a semi-infinite neutral sheet, by contrast, involves the creation of multiple x-type neutral points and does require considerable flux cancellation.

4.3 Efficiency of reconnection

Magnetic flux is advected toward the current sheet at a velocity

$$\frac{V}{c}\simeq\left(\frac{\mathrm{d}\,\tau_{\rm es}}{\mathrm{d}\ln r}\right)^{-1/2}, \tag{20}$$

when the scattering optical depth exceeds unity. This velocity is essentially the diffusion velocity of radiation generated in the neutral sheet. The fraction of the Poynting flux converted to gamma-rays depends directly on the kinematics as well as the opacity. In the wind rest frame, the distance between successive neutral sheets grows as $\Delta r/r_{\rm lc}\sim\pi\gamma$, and only a fraction $\simeq\pi^{-1}\gamma^{-2}(r/r_{\rm lc})$ of the magnetic flux remains in causal contact with a null surface. This fraction *decreases* with distance from the source while γ is growing, but begins to *increase* when γ has saturated at $\simeq\gamma_b$. The fraction of the Poynting flux that has annihilated is

$$f_{\rm ann}\simeq\frac{1}{\pi\gamma}\left(\frac{\mathrm{d}\,\tau_{\rm es}}{\mathrm{d}\ln r}\right)^{-1/2}$$
$$=8\times10^{-4}(Y_e L_{50}/P_{-3})^{-1/4}\left(\frac{\mathrm{d}\,\tau_{\rm es}}{\mathrm{d}\ln r}\right)^{-1/4}\left(\frac{\gamma_b}{\gamma_b^*}\right)^{1/3} \tag{21}$$

when $\gamma_b>\gamma_b^*$, and

$$f_{\rm ann}\simeq 8\times10^{-4}(Y_e L_{50}/P_{-3})^{-1/4}\left(\frac{\mathrm{d}\,\tau_{\rm es}}{\mathrm{d}\ln r}\right)^{-3/2}\left(\frac{\gamma_b}{\gamma_b^*}\right)^{-5} \tag{22}$$

where $\gamma_b<\gamma_b^*$. If a fraction ε_γ of the energy in the advected magnetic flux is transferred to the photons at a neutral sheet, then the change in photon luminosity is

$$\frac{\Delta L_\gamma}{L}=\varepsilon_\gamma f_{\rm ann}. \tag{23}$$

In deriving (21) and (22), we have assumed that the electron/baryon component is the dominant source of Thomson opacity. This is a reasonable approximation near radius r_τ for $\varepsilon_\gamma\lesssim 10^{-2}$ (see equation 16). When ε_γ is larger, one must multiply equations (21) and (22) by $0.26(\varepsilon_\gamma/Y_e)^{-1/4}$.

Note the strong dependence of $f_{\rm ann}$ on γ_b when the wind begins to coast inside the Thomson photosphere ($\gamma_b<\gamma_b^*$). There is a strong selection in favour of baryon loadings that are at least as heavy as

$$\gamma_b=\gamma_{b,\rm ann}\equiv 94(Y_e L_{50}/P_{-3})^{1/5}. \tag{24}$$

This is the value of γ_b for which $f_{\rm ann}=1$ at the Thomson photosphere. If the baryon loading is light ($\gamma_b\gg\gamma_{b,\rm ann}$), only a tiny fraction of the flux can annihilate inside the photosphere. The quantity $\gamma_{b,\rm ann}$ also represents the value of γ_b at which the minimum variability time-scale $t_{\rm var}\sim\gamma_b^{-2}(r_\tau/c)$ is comparable to the rotation period of the source, $t_{\rm var}\sim\pi r_{\rm lc}/c=P/2$.

5 NON-THERMAL GAMMA-RAY BURST SPECTRA

A key feature of 'classical' gamma-ray bursts[11] is their highly non-thermal spectra. A crude classification may be made in terms of two power laws joined at an energy E_0 which is broadly distributed but lies typically in the range $E_0\sim 100$–300 keV (Band et al. 1993). The spectral index[12] above this break is clustered around $\beta=-2$, which corresponds to constant flux per logarithmic photon energy. This refers to an average over the entire burst, but individual bursts often show considerable softening and steepening of the high-energy power law, with the burst duration sometimes measurably shorter at higher energies[13] (Band et al. 1992, 1993).

An important advantage of thermal cosmological fireballs is that they relate the observed gamma-ray fluence to the energy E_0 in a straightforward way (cf. Paczyński 1986; Goodman 1986). The photon temperature boosted to the observer's frame is related simply to $L_\gamma=\frac{4}{3}\gamma^2 aT_r^4\times4\pi r^2 c$ by

$$\frac{4}{3}\gamma T_r=0.67\,L_{\gamma,50}^{1/4}P_{-3}^{-1/2}\left(\frac{\gamma}{r/r_{\rm lc}}\right)^{1/2}\ \mathrm{MeV}, \tag{25}$$

when the photon distribution function in the wind rest frame is a blackbody at temperature T_r. This assumes that $kT_r\ll m_e c^2$, so that pairs can be neglected.

How can one retain this advantage of thermal fireballs, while at the same time obtaining a spectrum that is flat in energy above 1 MeV? We suggest that Compton scattering by Alfvén turbulence provides a simple solution.

Turbulence will be excited in an MHD wind by magnetic reconnection (Section 4) or by hydrodynamical instabilities such as the kink instability (e.g. Eichler 1993). It is visible in strongly collimated, radiative jets, and, in gamma-ray burst models, provides a means of accelerating energetic particles at large distances from the source (DT92). In the ultralumin-

[11] Given the lack of easily measured features by which to classify bursts, we define 'classical' bursts as all bursts except for the soft gamma repeaters.

[12] We use the notation of Bland et al. (1993) in which the spectral index is $\beta=\mathrm{d}\ln(\mathrm{d}N_\gamma/\mathrm{d}E_\gamma)/\mathrm{d}\ln E_\gamma$.

[13] It seems likely that the shorter duration of the high-energy emission is a direct consequence of a steepening of the spectrum, although to our knowledge the connection between these phenomena has not yet been quantified.

ous, relativistic winds under consideration here, this turbulence has some special properties.

First, Alfvén waves propagate at very nearly the speed of light. The baryon rest energy density is much lower than the magnetic energy density in the wind rest frame, $\rho_{b,r}c^2/(B_r^2/8\pi)=\gamma/\gamma_b$, so long as the bulk Lorentz factor has not yet reached the limiting value γ_b set by the baryon inertia. The phase velocity of the Alfvén waves corresponds to a high Lorentz factor $\gamma_A^2=\gamma_b/\gamma \gg 1$. (We are interested in linear waves whose transverse velocity is somewhat less than the speed of light.)

Secondly, the turbulence is damped effectively by Compton drag where the Thomson optical depth in the wind rest frame is greater than unity and the photon fluid carries a substantial fraction of the total energy density. This promises to be a more effective way of extracting energy from the magnetic field than electrostatic acceleration at current sheets (Section 4).

Incoherent Alfvén turbulence has a dramatic effect on the gamma-ray signal emerging from the wind. On scales smaller than the rest-frame horizon of the wind, bulk electron motions associated with the turbulence cause a Compton spectral distortion identical to that caused by hot thermal electrons.[14] The superposition of many waves results in a cancellation in the first-order Doppler shift. So long as the electrons are not fully relativistic, the dominant effect is diffusion in frequency space with a diffusion coefficient proportional to the scattering rate and to the mean-square frequency shift per scattering,

$$\frac{1}{3}n_e\sigma_T c\langle(\Delta\nu)^2\rangle=\frac{1}{3}n_e\sigma_T c\langle\gamma_w^2-1\rangle\nu^2.$$

Here, n_e is the rest-frame electron density, V_w is the transverse Alfvén wave velocity, $\gamma_w=(1-V_w^2/c^2)^{-1/2}$, and $\langle\ldots\rangle$ denotes a spatial averge. The equivalent electron temperature is

$$\begin{aligned}\frac{kT_w}{m_ec^2}&=\frac{1}{3}\langle\gamma_w^2-1\rangle\\&\simeq\frac{1}{3}\frac{\langle V_w^2\rangle}{c^2}\qquad(V_w/c\ll 1).\end{aligned}\tag{26}$$

When the turbulence is spatially bounded (as in the reconnecting wind model outlined in Section 4), the resulting photon spectrum in the wind rest frame is a power law extending from a lower energy E_{break} up to $E_\gamma\sim 4kT_w$. This property was first deduced by Shapiro, Lightman & Eardley (1976) in their treatment of Compton scattering by hot accretion disc coronae (see also Pozdnyakov, Sobol' & Sunyaev 1983). The spectrum below energy E_{break} is also a power law. The value of E_{break} lies close to the mean energy of the Comptonized photons, which are themselves generated closer to the source when thermalization is rapid. Notice that kT_w is never larger than $\sim\frac{1}{3}m_ec^2$, since $V_w/c\lesssim 1$.

The next two terms in the non-relativistic expansion of (26) are

$$\frac{kT_w}{m_ec^2}=\frac{1}{3}\frac{\langle V_w^2\rangle}{c^2}\left[1+\frac{5}{6}\frac{\langle V_w^2\rangle}{c^2}+\frac{5}{8}\left(\frac{\langle V_w^2\rangle}{c^2}\right)^2+\ldots\right]\tag{27}$$

[14]For a derivation in a cosmological context, see e.g. Peebles (1971).

when the electrons follow a (relativistic) Boltzmann distribution (Cooper 1971). We do not expect V_w^2 to follow a Boltzmann distribution in the case of Alfvén turbulence, so the relativistic corrections to (26) will differ from (27), but remarkably the spectrum turns out to be independent of these corrections. Additional relativistic corrections are present when $E_\gamma\sim m_ec^2$ (cf. Cooper 1971), but these are important only near the upper spectral cut-off, and only when $kT_w\sim m_ec^2$.

Waves of size comparable to the horizon generate a first-order Doppler shift of the spectrum. The sign of this shift is random, but is constant across the segment of the wind (of angular size $\sim\gamma^{-2}$) that is visible to the observer. This effect would be difficult to detect in practice, because most gamma-ray bursts do not have sharp, well-defined spectral breaks. It will provide an additional source of dispersion in the relation between total gamma-ray fluence and spectral break energy. Waves of size much larger than the horizon advect the photon fluid with few dissipative losses.

In this section, we calculate the high- and low-energy spectral indices β and α, and then the value of the spectral break energy E_{break}.

5.1 Photon escape from a bounded turbulent region

Fermi (1949) was the first to note that a power-law spectrum of particle energies could be generated if the particles (in our case, photons) were accelerated in a bounded volume, with the energy of each particle increasing at the rate $dE_\gamma/dt=\Gamma_{acc}E_\gamma$, and each particle having a probability Γ_{esc} per unit time of escaping from the acceleration volume. The resulting spectral index is

$$\beta=-1-\frac{\Gamma_{acc}}{\Gamma_{esc}}.\tag{28}$$

This mechanism was applied to the power-law X-ray spectrum of Cyg X-1 (the candidate black hole binary) by Shapiro et al. (1976), who showed that diffusion of soft photons out of a hot, homogeneous electron cloud results in a high-energy power-law tail, with spectral index

$$\beta=\frac{1}{2}-\left(\frac{9}{4}+y^{-1}\right)^{1/2}.\tag{29}$$

Here,

$$y=a(\Delta\tau_{es})^2\frac{kT_e}{m_ec^2}\tag{30}$$

is a suitably defined Compton parameter, which depends on the optical depth $\Delta\tau_{es}$ across the cloud. Each photon escaping from the cloud undergoes $\sim(\Delta\tau_{es})^2$ scatterings on average. The constant a is independent of the distribution of photon sources throughout the cloud, since the spectral index (29) applies to a small number of high-energy photons which remain trapped in the cloud for much longer than the mean residency time. This constant depends (weakly) on the geometry of the cloud; one finds $a=\pi^2/12$ for a bounded slab (Sunyaev & Titarchuk 1980).

Fermi's original acceleration model suffered from a basic defect: in order to reproduce the observed spectral index of cosmic rays, the acceleration time $\Gamma_{\rm acc}^{-1}$ and loss time $\Gamma_{\rm esc}^{-1}$ had to be nearly equal, and their ratio tuned to a specific value. This defect was removed by the mechanism of first-order shock acceleration, in which the ratio $\Gamma_{\rm acc}/\Gamma_{\rm esc}$ is fixed by the density compression ratio across the shock (e.g. Blandford & Eichler 1987).

The Compton cloud model for power-law X-ray spectra suffers from the same defect: the y-parameter must be adjusted to a specific value. A spectral index $\beta=-2$ requires $y=\frac{1}{4}$ (Shapiro et al. 1976).

This defect is largely resolved if the random electron motions are due to Alfvén turbulence. The amplitude of the turbulence is itself limited by Compton drag (cf. equation 33) in just the required manner. The photon energy density evolves according to

$$\frac{{\rm d}U_\gamma}{{\rm d}t}=4\,\frac{kT_{\rm w}}{m_{\rm e}c^2}\,n_{\rm e}\sigma_{\rm T}cU_\gamma-\tilde{\Gamma}_{\rm esc}U_\gamma, \tag{31}$$

where the second term represents escape of the photons from the turbulent region. In a steady state, the Compton parameter is

$$y=\tilde{a}N_{\rm scat}\,\frac{kT_{\rm w}}{m_{\rm e}c^2}\simeq\frac{1}{4},$$

where $N_{\rm scat}=n_{\rm e}\sigma_{\rm T}c\tilde{\Gamma}_{\rm esc}^{-1}$ is the mean number of times that a photon scatters in leaving the cloud. More remarkably, the high-energy photon spectral index is *attracted* towards $\beta=-2$ in an expanding, turbulent relativistic wind (Section 5.4).

It would be very difficult to achieve a similar regulation of y with thermal electron motions, or via electrostatic acceleration. In the absence of a magnetic field, the electrons could be shock-heated to a temperature $kT\sim m_{\rm e}c^2$, but there is no obvious mechanism that fixes the shock velocity at the required value. Moreover, most of the shock energy would be converted to random motions of the *baryonic* component of the plasma, which does not couple effectively to the electrons by Coulomb collisions near the Thomson photosphere. (By contrast, energy stored in Alfvén turbulence is transferred very effectively to the electrons.) Shocks may therefore be suppressed by particle heating near the photosphere.

The same regulation of y is required by Comptonization models for the power-law X-ray emission of Seyfert galaxies. The rapid variability of this emission (a property that is shared by the hard-spectrum 'low state' of Cyg X-1) is also consistent with Comptonization by Alfvén turbulence. Moreover, it has recently been suggested that the spectral index $\alpha_{\rm X}=-0.7$ ($\beta=-1.7$ in our notation) characteristic of Seyfert 1 galaxies is actually a superposition of a steeper intrinsic spectrum with index $\alpha_{\rm X}=-1$ ($\beta=-2$) and a flatter, reflected component (see Mushotsky et al. 1993, and references therein). These facts, taken together, lead us to propose that Comptonization by Alfvén turbulence is the source of this X-ray emission (Thompson 1994, in preparation).

This Compton heating mechanism can be applied directly to an expanding relativistic wind, even at high optical depth. The Alfvén turbulence generated by opacity-limited reconnection at a neutral sheet (Section 4.2) is confined to a layer of thickness $\Delta/ct_{\rm r}\sim({\rm d}\tau_{\rm es}/{\rm d}\ln r)^{-1/2}$, where $\tau_{\rm es}$ is the optical depth across the (rest-frame) horizon. Photons escape the layer in a time $\sim t_{\rm r}$.

The standard escape-probability formalism does, however, suffer from a basic difficulty when applied to gamma-ray bursts. The low-energy index is much too positive. For example, when the photon source inside the scattering cloud is effectively monochromatic, the shift in the spectral index across the break is $\alpha-\beta=3$ (e.g. Sunyaev & Titarchuk 1980), as compared with the observed break of $\alpha-\beta\simeq1$–2 (Band et al. 1993). A solution to this problem is found in Section 5.4, where we construct a self-similar solution for Compton scattering in a turbulent relativistic wind. When the seed photons are very soft, the resulting low-energy spectral index is $\alpha=-1$.

5.2 Alfvén turbulence in a relativistic magnetized wind

The spectrum arising from the interaction of a photon gas with Alfvén turbulence in a relativistic wind depends sensitively on the turbulent energy density. As a first step to calculating this spectrum, we write down the basic equations describing the evolution of the energy in waves and in photons.

The energy density in Alfvén waves with rms transverse velocity $V_{\rm w}$ is

$$U_{\rm w}\simeq\frac{\langle V_{\rm w}^2\rangle}{c^2}\frac{B_{\rm r}^2}{8\pi}, \tag{32}$$

when the baryon rest energy density is much less than the magnetic energy density in the wind rest frame, $\rho_{\rm b,r}c^2\ll B_{\rm r}^2/8\pi$. The waves lose energy to adiabatic expansion and to Compton drag off the ambient radiation field, and gain energy from large-scale instabilities (such as magnetic reconnection). The net gradient in the wave energy density $U_{\rm w}$ (as measured in the wind rest frame) is

$$\frac{{\rm d}U_{\rm w}}{{\rm d}r}=-\frac{\Gamma_{\rm w}\delta_\rho}{r}U_{\rm w}-\frac{4y(r)}{r}U_\gamma+\frac{\varepsilon_{\rm w}(r)}{r}\frac{B_{\rm r}^2}{8\pi}, \tag{33}$$

where the Compton parameter is defined by[15]

$$y\equiv\frac{1}{3}\langle\gamma_{\rm w}^2-1\rangle n_{\rm e}\sigma_{\rm T}c\frac{{\rm d}t_{\rm r}}{{\rm d}\ln r}=\frac{{\rm d}\tau_{\rm es}}{{\rm d}\ln r}\left(\frac{kT_{\rm w}}{m_{\rm e}c^2}\right). \tag{34}$$

Also, $\Gamma_{\rm w}$ is the adiabatic index of the Alfvén turbulence. The rest-frame baryon density scales with radius as

$$\frac{{\rm d}\ln\rho_{\rm b,r}}{{\rm d}r}=-\frac{\delta_\rho}{r}, \tag{35}$$

where $\delta_\rho=3$ when the wind is in free expansion, and $\delta_\rho=2$ when the bulk Lorentz factor has saturated at $\gamma\simeq\gamma_{\rm b}$. The rest-frame photon energy density U_γ decreases via adiabatic expansion, and increases via Compton drag,

$$\frac{{\rm d}U_\gamma}{{\rm d}r}=-\frac{4}{3}\frac{\delta_\rho}{r}U_\gamma+\frac{4y(r)}{r}U_\gamma. \tag{36}$$

[15]This definition differs from the usual one in that y is the Compton parameter accumulated per logarithm of radius. Note also the factor of 4 difference between our normalization (which follows the convention used in cosmological applications, e.g. Sunyaev & Zel'dovich 1980) and the normalization used by Shapiro et al. (1976).

Frictional damping occurs primarily at wavelengths smaller than $\sim[n_e\sigma_T(V_w/c)]^{-1}$, since the radiation is convected with the electrons at longer wavelengths. Thus damping occurs primarily on scales much smaller than the rest-frame horizon of the blast when the optical depth $d\tau_{es}/d\ln r \gg 1$.

The scaling of magnetic pressure with radius,

$$\frac{d\ln(B_r^2/8\pi)}{dr} = -\frac{\delta_B}{r}, \tag{37}$$

depends on the configuration of the field. If the field is kept tangled by stirring motions of some kind, then one always has

$$\delta_B = \frac{4}{3}\delta_\rho. \tag{38}$$

If the field is ordered and predominantly toroidal (as expected in the wind from a rotating source), then (38) still holds so long as the wind is in free expansion ($\gamma \propto r$), because in this case $\delta_B = 4$ and $\delta_\rho = 3$. The relation between δ_B and δ_ρ is different, however, when γ is limited by the baryon inertia. In this case, the phase coherence of the toroidal field with respect to the rotation of the source imposes $B \propto r^{-1}$, and so

$$\delta_B = \delta_\rho = 2. \tag{39}$$

Note also that, in a coasting phase (γ = constant), the radial component of the field scales as $\sim r^{-2}$ whereas the toroidal component scales as $\sim r^{-1}$, so that a tangled field is stretched primarily in the non-radial direction.[16] This suggests that the scaling $B \sim r^{-4/3}$ for a tangled field breaks down when the radial coherence length of the field is larger than the rest-frame horizon of the blast. Continual retangling of the field is prevented by causality, and the field is mainly toroidal outside a short distance from the source. For example, the scale of field variation in the wind from a rotating source is $\sim \gamma cP/2$ in the rest frame, whereas the rest-frame horizon is much smaller, $c(dt_r/d\ln r) = r/\gamma \sim cP/2\pi$. A value of δ_B intermediate between 2 and 8/3 is none the less possible if there is some continuous conversion between radial field and toroidal field.

The adiabatic index Γ_w of the Alfvén turbulence also varies with radius. The turbulence behaves like a relativistic fluid, $\Gamma_w = 4/3$, while the wind is still in free expansion and $\rho_{b,r}c^2 \ll B_r^2/8\pi$. In a coasting phase, Γ_w is intermediate between 4/3 and 5/3 so long as the magnetic field stays mainly toroidal and $\rho_{b,r}c^2 \sim B_r^2/8\pi$. Eventually the magnetic field tangles up, $B_r^2/8\pi$ drops well below $\rho_{b,r}c^2$, and Γ_w approaches 5/3.

A tangled field decreases faster with radius than a toroidal field in the coasting phase. This strongly suggests that the toroidal field can become unstable to tangling once the growth of γ is limited by the baryon inertia. None the less, the predominantly toroidal character of the field is protected by causality well beyond the radius where γ saturates. In the case of a rotating source, the field becomes fully tangled only outside the radius where $c(dt_r/d\ln r) \sim \gamma cP/2$, or equivalently where $r/r_{lc} \sim \pi\gamma^2$.

[16]This compares with the scaling $\sim r^{-2}$ for both rest-frame components of the field during free expansion.

5.2.1 *Implications for pulsar winds*

The implications of such an instability for the field geometry in pulsar winds can be briefly stated, as follows. A toroidal field carried by the wind of the Crab pulsar ($P = 33$ ms) could become tangled inside the radius where the wind interacts with the nebula (the 'wisp' radius $R \sim 3\times10^{17}$ cm: Rees & Gunn 1974). This would require that the inertia of the charged particles limit the bulk Lorentz factor to $\gamma \lesssim 3\times10^4$, assuming spherical symmetry. This Lorentz factor is lower than the value of $\sim 10^6$ often quoted for an electron–positron wind. None the less, it is compatible with limits set by induced Compton scattering of the radio pulses ($\gamma \gtrsim 1\times10^4$: Wilson & Rees 1978). Notice, however, that the tangling instability would also be triggered after the pulsar wind is shocked and decelerated by its interaction with the nebula. When the flow crosses the shock, the rest-frame horizon suddenly becomes much larger than the radial coherence length of the toroidal field. The post-shock Lorentz factor γ_{ps} is related to the upstream ratio σ of Poynting flux to particle energy flux by $\gamma_{ps} \simeq \sigma^{1/2}$, when $\sigma \gtrsim 1$ (Kennel & Coroniti 1984a). Tangling of the field would eliminate the need for a very small value of σ at the wisp radius, as suggested by a model of the post-shock nebular flow in which the magnetic field remains toroidal ($\sigma \sim 0.003$: Kennel & Coroniti 1984a). We are presently investigating the plasma heating produced by this instability, which will enhance the post-shock synchrotron emissivity and increase the amount of Poynting flux that is consistent with the observed synchrotron spectrum of the nebula (Kennel & Coroniti 1984b; De Jager & Harding 1992). By contrast, the magnetic field carried in the wind of the eclipsing millisecond pulsar 1957 + 20 ($P = 1.6$ ms) probably does not tangle inside the orbit of the companion ($a = 1.7\times10^{11}$ cm: Fruchter et al. 1990). This requires $\gamma_b \lesssim 80$, which is almost certainly much lower than the actual value. Randomization of the field in the shocked pulsar wind close to the companion remains a possibility, and may be testable by radio polarization measurements (Thompson et al. 1994).

5.3 Equilibrium wave density

The equilibrium wave energy density can now be calculated. When the optical depth is high, the waves rapidly lose energy to the photons, and one has $U_w/U_\gamma \ll 1$. The approximate solution to (33) is then

$$y(r) = \frac{\varepsilon_w(r)}{4\varepsilon_\gamma(r)}, \tag{40}$$

where $\varepsilon_\gamma \equiv U_\gamma/(B_r^2/8\pi)$.

Expression (40) for the Compton parameter is valid only so long as the optical depth $d\tau_{es}/d\ln r$ exceeds $\sim \varepsilon_w^{-1}$, and damping by Compton drag is faster than by adiabatic expansion. It has the important property that the equilibrium wave energy density is *independent* of the wave injection rate when (i) the waves are injected smoothly ($\varepsilon_\gamma \sim \varepsilon_w$) and (ii) the injection rate is high enough to maintain this equilibrium. We should emphasize that *the equilibrium value of* $\langle\gamma_w^2 - 1\rangle$ *would not be modified if it included a contribution from electron thermal motions in addition to Alfvén waves.*

While the wind is still in free expansion, the adiabatic indices of the Alfvén waves and photons are equal. Adding these equations and taking the wave generation rate to vanish inside radius r_w, one finds that the total energy in waves and photons is

$$\frac{U_w+U_\gamma}{B_r^2/8\pi}=\int_{r_w}^{r}\varepsilon_w(r')\frac{dr'}{r'}\equiv E_w(r). \tag{41}$$

In the particular case where ε_w is constant at $r>r_w$, $\varepsilon_w(r)=\varepsilon_{w0}\Theta(r-r_w)$, one finds

$$E_w(r)=\varepsilon_{w0}\ln(r/r_w). \tag{42}$$

This implies

$$y(r)\simeq\frac{1}{4}\ln(r/r_w) \tag{43}$$

at high optical depth.

Next let us suppose that wave generation begins only after the wind enters a coasting phase ($\gamma\simeq\gamma_b$). Well inside the photosphere, the Compton parameter is independent of radius, and

$$U_\gamma(r)=U_\gamma(r_w)\left(\frac{r}{r_w}\right)^{4y-(4/3)\delta_\rho}. \tag{44}$$

This implies a scaling solution $[U_\gamma(r)/B_r^2(r)=\text{constant}]$,

$$y=\frac{1}{3}\delta_\rho-\frac{1}{4}\delta_B \tag{45}$$

when the wave injection coefficient ε_w is constant. Larger values of the Compton parameter can be achieved after U_γ has been diluted by adiabatic expansion (outside a radius $\sim\gamma_b r_{lc}$), but before a large fraction of the magnetic energy has been transferred to the photons.

5.4 Spectral indices

We now calculate a self-similar solution for the photon occupation number, representing Compton scattering by Alfvén turbulence on scales smaller than the rest-frame horizon of the wind.

We start with the diffusion equation for photons undergoing Compton scattering [the Kompane'ets (1957) equation],

$$\frac{\partial n}{\partial t_r}+\dot{E}_\gamma\frac{\partial n}{\partial E_\gamma}=n_e\sigma_T c\frac{1}{E_\gamma^2}\frac{\partial}{\partial E_\gamma}\left[\frac{E_\gamma^4}{m_e c^2}\left(n+n^2+kT_w\frac{\partial n}{\partial E_\gamma}\right)\right]. \tag{46}$$

Here, n is the photon occupation number, and t_r is the time lapsed in the wind rest frame. The term in (46) proportional to $\dot{E}_\gamma$ represents the redshifting of photons in the rest frame. While the wind is still in free expansion ($\gamma\simeq r/r_{lc}$), one has

$$\frac{ct_r}{r_{lc}}=\ln(r/r_{lc}), \tag{47}$$

whereas

$$ct_r\simeq\frac{r}{\gamma_b} \tag{48}$$

when the bulk Lorentz factor is limited by the baryon inertia ($\gamma\simeq\gamma_b$). The rest-frame time coordinate increases very gradually before γ saturates.

The redshift term in (46) can be rewritten in such a way as to account for turbulent motions on the scale of the rest-frame horizon of the wind. One has

$$\dot{E}_\gamma(\partial n/\partial E_\gamma)=-[\tfrac{1}{3}\delta_\rho(\mathrm{d}\ln r/\mathrm{d}t_r)+\tfrac{1}{3}(\nabla\cdot\boldsymbol{v})]\partial n/\partial\ln E_\gamma,$$

with $\boldsymbol{v}$ being the turbulent peculiar velocity measured in the local wind rest frame. First-order Fermi acceleration will occur in regions of converging flow (as well as deceleration in regions of diverging flow). However, we do not expect first-order acceleration to be efficient even near the Thomson photosphere, since it requires moderately supersonic bulk fluid motions (Blandford & Payne 1981).

We look for a self-similar solution of (46) in the form of a broken power law,

$$n_<(t_r,E_\gamma)=n_{break}(t_r)\left[\frac{E_\gamma}{E_{break}(t_r)}\right]^{-2+\alpha}\qquad(E_\gamma<E_{break});$$
$$n_>(t_r,E_\gamma)=n_{break}(t_r)\left[\frac{E_\gamma}{E_{break}(t_r)}\right]^{-2+\beta}\qquad(E_\gamma>E_{break}). \tag{49}$$

We neglect both the recoil term and the stimulated term in (46), which is a valid approximation when $kT_w\gg E_{break}$ and $(kT_w/E_\gamma)\gg n$. The break energy $E_{break}(t_r)$ and normalization $n_{break}(t_r)$ are constrained by conservation of photon number, which implies

$$n_{break}(r)E_{break}^3(r)\propto\rho_{b,r}(r)\propto r^{-\delta_\rho}, \tag{50}$$

and conservation of energy, which implies

$$n_{break}(r)E_{break}^4(r)\propto U_\gamma(r). \tag{51}$$

Next let us substitute the ansatz (49) for $n_>$ in (46). We obtain

$$(1+\beta)\left(\frac{\mathrm{d}\ln U_\gamma}{\mathrm{d}\ln r}+\frac{4}{3}\delta_\rho\right)=(1+\beta)(2-\beta)n_e\sigma_T c\frac{\mathrm{d}t_r}{\mathrm{d}\ln r}\frac{kT_w}{m_e c^2}. \tag{52}$$

The low-energy spectral index α satisfies the same equation. Substitution of equation (36) into the left-hand side yields

$$(1+\beta)4y=(1+\beta)(2-\beta)y. \tag{53}$$

To choose between the two roots of equation (52), we require that the occupation number increases with time on the high-energy branch of the spectrum, and decreases with time on the low-energy branch. This fixes

$$\alpha=-1 \tag{54}$$

and

$$\beta=-2. \tag{55}$$

In our notation, β is the slope of $\mathrm{d}\ln N_\gamma/\mathrm{d}\ln E_\gamma$, so that $\beta=-2$ corresponds to constant energy flux per logarithm of photon energy. The scattering term and the time derivative in (46) both vanish for $\alpha=-1$; this spectral index corresponds to constant photon number density per logarithm of photon energy.

The spectral indices (54) and (55) are our first main result. Band et al. (1993) find that, out of 54 GRBs with spectra measured by BATSE, 38 bursts have high-energy indices in the range -2 ± 0.6 when the spectrum is averaged over the entire burst. Steeper spectra are measured for the remaining bursts, but this may be partly a result of low signal-to-noise ratio. (In our model, steeper spectra do arise when the turbulence is generated at low scattering depth; Section 5.7.) The low-energy spectral index lies typically in the range $\alpha=-1$ to 0 (Band et al. 1993).

We can also show that this solution for the spectral indices is an *attractor*. Taking α and β to be functions of radius and substituting the same ansatz for n into the Kompane'ets equation, one finds

$$\frac{d\beta}{d\ln r}=\frac{y(1+\beta)(2+\beta)}{\ln(E_\gamma/E_{\rm break})}, \tag{56}$$

with a similar expression for α. Writing $\alpha=-1+\Delta\alpha$ and $\beta=-2+\Delta\beta$, one easily sees that $\Delta\alpha\to0$ below the spectral break, and $\Delta\beta\to0$ above the spectral break.

Note that relativistic corrections to the Kompane'ets equations, as embodied in equation (26), cancel out. Although the photons diffuse through frequency space faster than one would deduce from the non-relativistic approximation $kT_{\rm w}/m_{\rm e}=\frac{1}{3}\langle V_{\rm w}^2\rangle$, the amplitude of the turbulence is reduced by precisely the factor required to compensate for this faster diffusion. Additional relativistic corrections to Thomson scattering enter when $E_\gamma\sim m_{\rm e}c^2$ (Cooper 1971), but photons with $E_\gamma\sim1$ MeV in the observer's frame have $E_\gamma\ll m_{\rm e}c^2$ in the blast rest frame, and so these corrections can be neglected. They are important only at high energies where spectra are not yet accurately measured. In addition, when $\beta=-2$, the photon occupation number at energy $E_{\rm break}$ is suppressed by a factor $\sim[\ln(kT_{\rm w}/E_{\rm break})]^{-1}$ which grows with radius; but the resulting correction to y can be shown to be small, of order $E_{\rm break}/kT_{\rm w}$.

Finally, we note that an increase in the photon number flux causes a hardening of the high-energy spectrum. We assume the scaling $n_\gamma\propto r^{-\delta_\gamma}$ for the photon density in the wind rest frame, with $\delta_\gamma<\delta_\rho$. The corresponding solution is

$$\beta=-2+\frac{1}{3y}(\delta_\rho-\delta_\gamma). \tag{57}$$

This reduces to $\beta=-2+2(\delta_\rho-\delta_\gamma)$ during a coasting phase ($\gamma=\gamma_{\rm b}$). The strong dependence of the spectral index on $\delta_\rho-\delta_\gamma$ shows that $\beta=-2$ is characteristic of Comptonization by Alfvén turbulence, *only to the extent that photon number is conserved during the scattering process*. Photon number-changing reactions are probably inefficient near the Thomson photosphere (see Section 5.6). A relationship similar to (25) between spectral break energy and total gamma-ray energy flux requires that *the photon luminosity be comparable to the total spin-down luminosity at the radius where photon number-changing processes freeze out.*

5.5 Spectral breaks and cut-offs

We have calculated the equilibrium photon spectrum resulting from scattering by mildly relativistic Alfvén turbulence. This spectrum is a broken power law with index $\alpha\simeq-1$ below photon energy $E_{\rm break}$, and index $\beta\simeq-2$ between $E_{\rm break}$ and an upper cut-off $E_{\rm max}\sim4kT_{\rm w}$ in the blast rest frame. We identify the energy $E_{\rm break}$ with the spectral break observed in (most) classical GRBs.

How is this non-thermal photon spectrum established? Note that the spectral indices are attracted towards the equilibrium values (54) and (55) only rather slowly when $y\sim1/4$, unless E_γ is close to $E_{\rm break}$. A turbulent, relativistic wind will *maintain* a high-energy spectral index $\beta=-2$; but the *generation* of this power law probably requires that the turbulence be spatially bounded, as in the reconnection model outlined in Section 4. A low-energy spectral index $\alpha=-1$ develops as the seed photons (whose energy density has been diluted by adiabatic expansion outside a radius $\sim\gamma_{\rm b}r_{\rm lc}$) are upscattered until they have soaked up almost all the turbulent wave energy ($\varepsilon_\gamma\sim\varepsilon_{\rm w}$). That is, this low-energy power law is cut off at an energy $E_{\rm min}\sim E_{\rm seed}$. (When $E_{\rm seed}$ is very small, stimulated scattering effects cut off the spectrum at a higher energy; see below.)

Very close to the source, thermalization is rapid and the spectrum is very nearly blackbody. There is a characteristic radius $r_{\rm eq}$ where $E_{\rm break}\sim E_{\rm max}\sim E_{\rm min}$, namely

$$\frac{r_{\rm eq}}{r_{\rm lc}}\simeq1.4\times10^2(Y_{\rm e}L_{50}/P_{-3})^{3/16}\left(\frac{\gamma_{\rm b}}{\gamma_{\rm b}^*}\right)^{-1/4}\left(\frac{\varepsilon_\gamma}{\varepsilon_{\rm w}}\right)^{1/4}\times\left[\frac{E_{\rm break}(\text{observed})}{1\ \text{MeV}}\right]^{1/4}. \tag{58}$$

Inside this radius, the Compton recoil [the term in (46) proportional to n] cannot be neglected. The photons follow a Wien distribution, and the photon temperature is very nearly equal to the electron temperature. The y-parameter rises above the equilibrium value (40) toward smaller radii.

Outside the radius $r_{\rm eq}$, the mean energy of the electrons rises above the mean energy of the photons. It is not energetically possible for all the photons to be upscattered to an energy comparable to $kT_{\rm w}$, and so the mean photon energy remains close to $E_{\rm break}$. The power-law segments of the spectrum grow wider as the ratio of $kT_{\rm w}$ to $E_{\rm break}$ grows. In the observer's frame, the upper spectral cut-off grows with radius, whereas $E_{\rm break}$ remains constant, and $E_{\rm min}$ declines. The slope of each power-law segment of the spectrum remains constant, so long as the injection rate of turbulent energy is sufficient to maintain it, and photon number is conserved (Section 5.4). Photon number-changing processes include double Compton scattering $e+\gamma\to e+\gamma+\gamma$ (Lightman 1981), bremsstrahlung and synchrotron emission from a small contaminant of relativistic particles. It is straightforward to show that double Compton emission dominates bremsstrahlung as a source of soft photons when $kT_{\rm w}\sim m_{\rm e}c^2$, but also that neither process is a significant photon source even at the lower cut-off energy $E_{\rm min}$, when $E_{\rm min}\ll kT_{\rm w}$.

Let us first discuss the high-energy spectral break. In the rest frame of the blast,

$$E_{\rm max}\sim4kT_{\rm w}=m_{\rm e}c^2\left[\frac{\varepsilon_\gamma(r)}{\varepsilon_{\rm w}(r)}\frac{d\tau_{\rm es}}{d\ln r}\right]^{-1}. \tag{59}$$

If the wind is in a coasting phase ($\gamma\simeq\gamma_{\rm b}$), then

$$E_{\rm max}\propto r\qquad(r_{\rm eq}<r<r_\tau). \tag{60}$$

Near the photosphere, the upper cut-off saturates at an energy $E_{\rm max}\sim m_{\rm e}c^2$, which corresponds to $E_{\rm max}(\text{observed})\sim\frac{4}{3}\gamma m_{\rm e}c^2$ in the observer's frame. For low baryon loadings ($\gamma_{\rm b}\sim\gamma_{\rm b}^*$) this cut-off is slightly smaller than the maximum photon energy $E_{\rm max}\sim1$ GeV measured by *Comptel* from

GRB 930131 (Sommer et al. 1994). From equation (13), one deduces

$$E_{\rm max}({\rm observed}) \sim 300 (L_{\gamma,50}/P_{-3})^{1/4}\ {\rm MeV}. \tag{61}$$

The maximum Lorentz factor γ_b^* attainable inside the Thomson photosphere (and therefore the cut-off energy $E_{\rm max}$) would increase if the wind were collimated about the rotation axis. Note also that pair production by photon collisions can increase γ_b^* by as much as a factor ~ 4 (Section 3.3).

The parameters $E_{\rm break}$ and $n_{\rm break}$ are related to the photon number density

$$n_\gamma = \frac{E_{\rm break}^3 n_{\rm break}}{\pi^2 \hbar^3 c^3}\left[1 + \ln\left(\frac{E_{\rm break}}{E_{\rm min}}\right)\right], \tag{62}$$

and the total photon energy density

$$U_\gamma = \frac{E_{\rm break}^4 n_{\rm break}}{\pi^2 \hbar^3 c^3}\left[1 + \ln\left(\frac{E_{\rm max}}{E_{\rm break}}\right)\right], \tag{63}$$

in a straightforward manner. Whatever process determines the total photon density also determines the low-energy cut-off $E_{\rm min}$, since $n_\gamma \sim 3E_{\rm min}^2 kT_{\rm w}/(\pi^2\hbar^3c^3)$ (neglecting logarithmic factors). At a fixed comoving value of the photon energy, $\tilde{E}_\gamma \propto r^{\delta_\rho/3} E_\gamma$, the occupation number grows as $n(\tilde{E}_\gamma) \propto E_{\rm w}(r) r^{4/3\delta_\rho - \delta_B}$ on the high-energy segment; whereas it is slowly suppressed, $n(\tilde{E}_\gamma) \propto [1 + \ln(E_{\rm break}/E_{\rm min})]^{-1}$, on the low-energy segment.

The spectral break energy may be written as

$$E_{\rm break} = \frac{U_\gamma}{n_\gamma}\frac{1 + \ln(E_{\rm break}/E_{\rm min})}{1 + \ln(E_{\rm max}/E_{\rm break})}. \tag{64}$$

The scaling of $E_{\rm break}$ with radius differs in one dramatic respect from the scaling of the peak energy of a thermal fireball. From equations (50) and (51), the observed break energy

$$E_{\rm break}({\rm observed}) = \frac{4}{3}\gamma E_{\rm break} \tag{65}$$

scales as

$$E_{\rm break}({\rm observed}) \propto \gamma(r)\,\varepsilon_\gamma r^{\delta_\rho - \delta_B}. \tag{66}$$

So long as the magnetic field remains almost toroidal (which we have argued is the case even if the field starts off tangled close to the source), this radial variation of the break energy is quite weak:

$$E_{\rm break}({\rm observed}) \propto \varepsilon_\gamma. \tag{67}$$

This relation holds *both* in the free-expansion phase ($\gamma \propto r$, $\delta_\rho = 3$, $\delta_B = 4$) and in the coasting phase ($\gamma = $ constant, $\delta_\rho = \delta_B = 2$). By contrast, the peak energy of the thermal fireball is constant in the free-expansion phase, but decays as $r^{-2/3}$ in the coasting phase. The main difference here is that the Alfvén turbulence constantly upscatters the radiation, in such a way as to compensate adiabatic softening during the coasting phase. We conclude that *the observed spectral break is not substantially softened by heavy baryon loading*, so long as the field does not become highly tangled well inside the Thomson photosphere. We gave the corresponding lower bound on γ_b for a spherical wind in Section 4.3 (equation 24).

Further insight into the evolution of the photon spectrum is provided by an integral solution to the Kompane'ets equation. If we neglect the stimulated term and recoil term in (46), the remaining Doppler term can be re-expressed as a diffusion term (Sunyaev & Zel'dovich 1980). The solution (modified to account for the effects of adiabatic expansion) can be written in terms of the initial occupation number $n_0(E_\gamma)$ as

$$n(E_\gamma, \Delta y) = \frac{1}{\sqrt{4\pi\Delta y}}\int_0^\infty n_0(E_{\gamma 0})\exp\left\{-\frac{1}{4\Delta y}\left[\ln(E_\gamma/E_{\gamma 0}) - 3\Delta y + \frac{1}{3}\delta_\rho \ln(r/r_0)\right]^2\right\}\frac{dE_{\gamma 0}}{E_{\gamma 0}}. \tag{68}$$

Here, $\Delta y = \int_{r_0}^r y(dr'/r')$ is the total Compton parameter accumulated from radius r_0 to radius r. The change in $E_{\rm break}$ can be approximately derived by setting the exponent in this equation to zero. This gives

$$\Delta y \simeq \frac{1}{8}\ln^2(r/r_0) + \frac{1}{4}\ln(r/r_0)\ln(r_{\rm w}/r_0);$$

$$\ln(E_{\rm break}/E_{\rm break,0}) \simeq -\ln(r/r_0) + \frac{3}{8}\ln^2(r/r_0) + \frac{3}{4}\ln(r/r_0)\ln(r_{\rm w}/r_0) \tag{69}$$

during free expansion, and

$$\Delta y \simeq \frac{1}{6}\ln(r/r_0);$$

$$\ln(E_{\rm break}/E_{\rm break,0}) \simeq -\frac{1}{6}\ln(r/r_0) \tag{70}$$

during a coasting phase. The accumulated Compton parameter grows faster with radius during a free-expansion phase, which would indicate that the photons spread out faster in frequency during such a phase. This effect is marginal in practice, however, because the photons remain thermal out to the relatively large radius $r_{\rm eq}$ (equation 58), and the range of radii between $r_{\rm eq}$ and the end of free expansion ($r \sim \gamma_b r_{\rm lc}$) is limited. This also shows that the break energy declines below the scaling (66)–(67) as the amplitude of the Alfvén turbulence declines below the equilibrium value (45).

The numerical value of $E_{\rm break}$ depends on the number of photons carried by the wind. If n_γ and U_γ are related in the same way as in a blackbody, then

$$E_{\rm break}({\rm observed}) \simeq 1.8\, L_{\gamma,50}^{1/4} P_{-3}^{-1/2}\left[\frac{1 + \ln(E_{\rm break}/E_{\rm min})}{1 + \ln(E_{\rm max}/E_{\rm break})}\right]{\rm MeV} \tag{71}$$

in the observer's frame (cf. equation 25). When the lower cut-off $E_{\rm min}$ is set by stimulated effects, $E_{\rm break}/E_{\rm min} \sim (kT_{\rm w}/E_{\rm break})^{1/2}$, and so the expression in brackets is approximately 1/2 when $kT_{\rm w} \gg E_{\rm break}$. As a result,

$$E_{\rm break}({\rm observed}) \simeq 0.9\, L_{\gamma,50}^{1/4} P_{-3}^{-1/2}\ {\rm MeV}. \tag{72}$$

This energy lies comfortably close to the observed spectral break energies (~ 100–300 keV with a tail extending above 1 MeV: Shaefer et al. 1992; Band et al. 1993) for a photon luminosity of order $\sim 10^{50}$ erg s^{-1}. Note that the break energy should depend weakly on the degree of collimation of the wind, given a fixed value of the gamma-ray flux reaching the observer.

Higher luminosities are predicted to be associated with higher break energies. The spectral break energy depends on other quantities such as the source size, cosmological redshift, and amount of baryon loading, so a strong correlation between $E_{\rm break}$ and flux is not expected between different bursts. However, such a correlation should be present *within* an individual burst, as is indeed seen in the particular case of GRB 930131 (Kouveliotou et al. 1994). $E_{\rm break}$ is a bit too high if the source region is much smaller than a neutron star ($P_{-3} \sim 0.1$ for a 2-$M_\odot$ extremal Kerr BH). Of course, a small amount of adiabatic softening inside the Thomson photosphere (Section 4.3) would compensate for the effects of a small source volume.

$E_{\rm break}$ is also reduced below the estimate (71) if new photons are created outside the radius $r/r_{\rm lc} \sim \gamma_{\rm b}$ (where the coasting phase begins). This could generate a gamma-ray spectrum that is essentially flat over the range of energies (from ~ 40 keV up to ~ 1–10 MeV) observed by the BATSE spectroscopic detectors. Perhaps 25 per cent of the bursts included in the most recent catalogue (Shaefer et al. 1993) have this property. None the less, photon creation must stop well inside the Thomson photosphere if the high-energy spectral index is to remain close to $\beta = -2$.

Now we apply these results to the reconnecting wind model outlined in Section 4. Consider the following situation, in which the photon energy flux L_γ is equal to the low-frequency Poynting flux L at radius $r_{\rm eq}$. First, we calculate the spectral break energy for $\gamma_{\rm b} = \gamma_{\rm b,\,ann}$. Out to a radius $\sim \gamma_{\rm b} r_{\rm lc}$, both photon energy flux and Poynting flux decrease at the same rate; in between this radius and the photosphere the Poynting flux remains constant, but both L_γ and the mean photon energy are reduced by a factor $\sim (r_\tau/\gamma_{\rm b} r_{\rm lc})^{-2/3}$ due to adiabatic expansion. We assume that the photon number flux remains constant. If most of the Poynting flux is then converted into gamma-rays near the photosphere (while conserving photon number), the photons are again upscattered in energy. The photon energy density is restored to the value that it would have had in the absence of baryon loading, and the spectral break energy is given precisely by equations (71) and (72). In this model, the gamma-ray luminosity is unaffected by baryon loading for values of $\gamma_{\rm b}$ greater than the critical value $\gamma_{\rm b,\,ann}$ given by equation (24). When $\gamma_{\rm b}$ is smaller, the magnetic field tangles up well inside the Thomson photosphere, and the emergent gamma-ray luminosity scales as $L_\gamma \propto (r_\tau/\pi\gamma_{\rm b}^2 r_{\rm lc})^{-2/3} \sim (\gamma_{\rm b}/\gamma_{\rm b,\,ann})^{10/3}$. Indeed, L_γ *decreases* with increasing neutrino luminosity: $L_\gamma \propto L_\nu^{-34/9}$, assuming the scaling (8). This suggests that L_γ is bounded above. None the less, spectral softening by adiabatic expansion requires a higher baryon loading than in a standard thermal fireball (where it occurs for $\gamma_{\rm b} < \gamma_{\rm b}^*$ rather than $\gamma_{\rm b} < \gamma_{\rm b,\,ann}$; e.g. Paczyński 1990).

The low-energy spectral index (54) holds only for photon energies where the stimulated term in (46) can be neglected, $(n_<)^2 \lesssim kT_{\rm w}|\partial n_</\partial E_\gamma| = 3(kT_{\rm w}/E_\gamma)\, n_<$. This power law is cut off from below at an energy

$$\frac{E_{\rm min}}{E_{\rm break}} = \left(n_{\rm break}\,\frac{E_{\rm break}}{3kT_{\rm w}}\right)^{1/2} \tag{73}$$

when the energy of the seed Comptonized photons is very small. This energy scales with radius as

$$E_{\rm min}(r) \propto r^{-3/2} \tag{74}$$

when $\gamma \simeq \gamma_{\rm b}$.

Finally, we should emphasize that the energy flux at much lower frequencies (roughly from soft X-rays to the radio) will be supplemented by secondary emission processes involving the interaction of the wind with ambient material (Paczyński & Rhoads 1993; Mészáros & Rees 1993). This material could derive from a companion star or disc, from a pre-supernova wind, or from the interstellar medium. Study of such secondary emission is outside the scope of this paper.

5.6 Spectral evolution near the Thomson photosphere

The photons decouple from the matter at the radius where $d\tau_{\rm es}/d\ln r = \frac{1}{4}(\varepsilon_{\rm w}/\varepsilon_\gamma)$:

$$r_\gamma = r_\tau \left(\frac{1}{4}\frac{\varepsilon_{\rm w}}{\varepsilon_\gamma}\right)^{-1/3}, \tag{75}$$

which is close to the Thomson photosphere (14). This corresponds to the bulk Lorentz factor

$$\gamma_\gamma = 3.9 \times 10^2 (Y_{\rm e} L_{50}/P_{-3})^{1/4} \left(\frac{\gamma_{\rm b}}{\gamma_{\rm b}^*}\right)^{-1/3} \left(\frac{1}{4}\frac{\varepsilon_{\rm w}}{\varepsilon_\gamma}\right)^{-1/3} \tag{76}$$

for $\gamma_{\rm b} > \gamma_{\rm b}^*$, with an analogous expression $\gamma_\gamma \propto (\gamma_{\rm b}/\gamma_{\rm b}^*)^{-3}$ for $\gamma_{\rm b} < \gamma_{\rm b}^*$. This Lorentz factor has been calculated assuming spherical symmetry, and the true value would be higher if the wind were collimated. The numerical value of the decoupling radius $r_\gamma \sim 10^9$ cm for $L_{50} \sim P_{-3} \sim 1$. Notice that r_γ lies far inside the radius of $\sim 10^{14}$ cm at which the gamma-rays are generated in the synchrotron self-Compton model of Mészáros & Rees (1993).

Collisions of photons of energy $E_\gamma \sim 3kT_{\rm w}$ can produce cold pairs[17] when $d\tau_{\rm es}/d\ln r \lesssim 1$. They have a higher density than the thermal freeze-out density (which has been calculated by Shemi & Piran 1990). These pairs also contribute more to the scattering depth that the electron–baryon contaminent if $\varepsilon_{\rm w} \sim 1$ [cf. (16)], but the resulting increase in the Thomson photosphere (the radius r_γ) is only a factor of $\simeq 3$. Note also that these pairs are far too cold to produce a synchrotron-pair cascade.

At low optical depths, the Alfvén wave amplitude saturates at $V_{\rm w} \sim c$, which corresponds to $y \sim (1/3)(d\tau_{\rm es}/d\ln r)$. The spectral indices (54) and (55) are preserved in the transition from high to low optical depth, because the expected frequency shift for any *single* high-energy photon is small near $d\tau_{\rm es}/d\ln r \sim 1$. The Compton recoil affects only photons with energies higher than the (rest-frame) high-energy cut-off $E_{\rm max} \sim 4kT_{\rm w}$. That is, significant downscattering of photons through the energy interval $E_{\rm break} < E_\gamma \lesssim 4kT_{\rm w}$ requires a much larger accumulated Compton parameter,

$$\Delta y = \int_{r_0}^{r_\gamma} y(r')\,\frac{dr'}{r'},$$

[17]This contrasts with models in which particle acceleration occurs so far from the source that γ–γ pair production can be neglected (e.g. Mészáros & Rees 1993).

than is provided by the turbulent motions near the photosphere. Starting from the estimate $3\Delta y - \frac{1}{3}\delta_\rho \ln(r/r_0) \gtrsim \ln(4kT_w/E_\gamma)$ (equation 68), one finds that relaxation of the high-energy spectrum above energy E_γ is possible only if the turbulence turns off inside a radius $r_0 \ll r_\gamma$, where

$$\frac{r_\gamma}{r_0} \sim \left(\frac{4kT_w}{E_\gamma}\right)^{1/(\delta_\rho/3-3y)} \gg 1 \qquad (77)$$

when y is independent of radius.

It is possible that the magnetic field stays predominantly toroidal out to the Thomson photosphere when γ_b lies somewhat below the critical value $\gamma_{b,ann}$. In this case, neither E_{break} nor L_γ is greatly reduced by adiabatic expansion. None the less, a heavy baryon loading still has two important observational consequences: the minimum variability time t_{var} is lengthened [in proportion to $(\gamma_b/\gamma_{b,ann})^{-5}$], and the high-energy spectral cut-off lies at a relatively low energy $(\sim\frac{4}{3}\gamma_b m_e c^2)$. This suggests a positive correlation between the *degree of variability* of a burst, and the position of the *high-energy spectral cut-off* (but not necessarily the position of the spectral break). For example, a baryon loading as heavy as $\gamma_b \sim 30$ will result in a very smooth burst which is cut off at ~ 30 MeV, but which could also have a break energy of order a few hundred keV.

5.7 Spectral variations

The wave energy density and the Compton parameter achieve the equilibrium value (40) at the Thomson photosphere only if the Alfvén turbulence is *strong*, that is, only if $\varepsilon_w \sim 1$. Lower injection rates of turbulent energy will result in steeper spectra. When $\varepsilon_w < 1$, the Compton parameter begins to decrease from the equilibrium value (40) after the wave temperature rises above $kT_w/m_e c^2 \sim \frac{1}{3}\varepsilon_w$, or equivalently after the optical depth drops below $d\tau_{es}/d \ln r \sim (2\varepsilon_w)^{-1}$. This implies that $y \sim \frac{1}{3}\varepsilon_w$ at the photosphere. The high-energy spectral index generated by photons escaping from a neutral sheet steepens to

$$\beta \sim \frac{1}{2} - \left(\frac{9}{4} + \frac{3}{\varepsilon_w}\right)^{1/2}.$$

In the wind reconnection model described in Section 4, most of the gamma-ray flux is generated near the Thomson photosphere when γ_b exceeds the critical value $\gamma_{b,ann}$. Thus steepening of the spectrum from $\beta = -2$ to $\beta = -3$ requires a reduction in the wave injection coefficient to $\varepsilon_w \sim 0.3$. However, when a $\beta = -2$ power law has been established and a significant fraction of the Poynting flux is converted to wave energy, very little steepening of the spectrum occurs (Section 5.6).

Some bursts do show clear evidence of spectral softening (Band et al. 1993). One clear case is GB830801b (Kuznetsov et al. 1986), where the high-energy tail remained a power law but steepened with time. A decrease of the high-energy index β from the beginning to the end of a burst does occur in this model if the injection amplitude ε_w of the Alfvén turbulence *decreases with time*. The low-energy spectral index, which according to (54) does not depend on y, is predicted to show less evolution.

Monte Carlo simulations of Comptonization of soft photons by warm electrons show that, when $\tau_{es} > 1$, the resulting power-law tail joins smoothly on to the soft photon peak *without* the appearance of a prominent, soft thermal bump. Such a bump appears only if the non-thermal tail is generated at $\tau_{es} < 1$ (e.g. fig. 3 of Pozdnyakov et al. 1983). It is certainly possible that the wave generation begins outside the scattering photosphere $(r_w > r_\tau)$, but we do not expect that generally to be the case.

6 CONCLUSIONS AND OPEN QUESTIONS

We have described a mechanism for producing a non-thermal X-ray and gamma-ray spectrum in ultraluminous and highly compact relativistic winds: Compton upscattering by mildly relativistic Alfvén turbulence. This pre-supposes an instability driven by magnetic reconnection or by cross-field stresses, but does not require any interaction of the wind with an external medium.[18] The spectrum is a broken power law, with high-energy index close to $\beta = -2$ $(\nu F_\nu = \text{constant})$ in the presence of large-amplitude turbulence, and low-energy index close to $\alpha = -1$. Steepening of the high-energy spectrum occurs if the injection rate of wave energy decreases, but the spectrum remains a power law. The low-energy spectral index is predicted to show less evolution.

These two power laws are joined at an energy which, when Lorentz-boosted into the observer's frame, lies near $1\,L_{\gamma,50}^{1/4}$ MeV if the source has a spin period of order 10^{-3} s. More compact sources produce a break at too high an energy unless either the spectrum undergoes a small amount of adiabatic softening inside the scattering photosphere, or the photon number flux is raised (perhaps by synchrotron emission from a small contaminant of relativistic particles). The upper cut-off lies at $E_\gamma \sim m_e c^2$ in the wind rest frame, which translates to an energy as high as $E_\gamma \sim 10^3 m_e c^2$ in the observer's frame, depending on baryon loading. Collimation of the wind allows a higher Lorentz factor inside the scattering photosphere, and hence a higher cut-off energy.

Perhaps our most remarkable finding is that the observed break energy is *not* greatly reduced by heavy mass loss, so long as the magnetic field stays mainly toroidal and conversion of Poynting flux to Alfvén turbulence continues right out to the Thomson photosphere. The observed gamma-ray spectrum emerges at a distance as small as $\sim 10^9$ cm from the source, depending on baryon loading. This is well outside the light cylinder, but also a factor $\lesssim 10^5$ inside the interaction radius assumed by the model of Mészáros & Rees (1993).

The model requires that a significant fraction of the Poynting flux must be converted to Alfvén turbulence and thence to gamma-rays in order to produce a high-energy spectral index $\beta = -2$. The total energy radiated over the duration of a burst is then $\sim 10^{51}(\Delta\Omega/4\pi)$ erg, much smaller than the binding energy of a neutron star. This suggests that the source retains a large store of rotational energy that could be tapped in subsequent outbursts if an appropriate trigger were available.

[18]The turbulence could be triggered when the wind is shocked and decelerated by an external medium (Section 5.2.1). Such an external trigger for the tangling instability is not needed in a cosmological GRB model, except when the baryon loading is extremely (unrealistically) low. Moreover, the large increase in the Thomson optical depth occurring behind the shock would probably cause excessive softening of the gamma-ray spectrum.

Perhaps the greatest strength of this model is that it depends only on generic properties of the source, namely a high temperature, rapid rotation, and a strong magnetic field (which is itself an inevitable consequence of the first two). A variety of rapidly rotating objects that have been suggested as sources of cosmological GRBs would produce a magnetized wind of the required luminosity (more than one of which may operate in nature): a white dwarf that undergoes accretion-induced collapse (directly to nuclear density) to form a neutron star with a millisecond spin period; or a Kerr black hole surrounded by a massive accretion disc (which is the possible end-product of a failed Type Ib supernova or of a binary neutron star merger). In all cases, the magnetic field required to generate the spin-down luminosity of $\sim 10^{50}$ erg s^{-1} (modulo beaming) is most plausibly generated by a *helical dynamo operating in a convective, differentially rotating body with a mean density close to that of nuclear matter* (DT92). The required poloidal field is $B \sim 5 \times 10^{14}$ $(M/2\ M_\odot)^{-1}$ G for material in Keplerian orbits about a central mass M. This amplification mechanism for a magnetic field is based on direct empirical knowledge of magnetic activity in stars (DT92; TD93). The same radiative mechanism, however, would probably also operate in an MHD wind powered by multiple disc flares (Narayan et al. 1992). The tendency of an advected magnetic flux rope to become toroidal outside the light cylinder of an expanding relativistic wind (Section 5.2) would allow a significant contribution to the total wind luminosity from flares with little change in the physical conditions at the scattering photosphere.

It has been observed that the durations of classical GRBs cover a range (~ 1–100 s) that is not too different from the neutrino cooling time (~ 3 s to release half the internal heat) of a newly formed neutron star (Dar et al. 1992). Although direct conversion of neutrinos to gamma-rays near the neutrinosphere (Goodman, Dar & Nussinov 1987) is problematic as a source of gamma-ray bursts, it should be emphasized that the ~ 30-s lifetime for the *convective motions* in a new-born neutron star is only slightly longer. (More precisely, this is the time for the star to become optically thin to neutrinos). When convection turns off, a large-scale dynamo-generated poloidal field will weaken, and the spin-down torque will drop off (TD93). The observed range of burst time-scales could easily be accommodated by a range in neutron star/disc mass and geometry, and thus in the convective lifetime.

If the Alfvén turbulence is excited by reconnection, then there is a strong selection in favour of burst sources with moderate baryon loadings, corresponding to limiting bulk Lorentz factors $\lesssim 300$ in a spherical wind. Only a small fraction of the magnetic flux carried by the wind is able to reconnect inside the Thomson photosphere when the mass-loss rate is lower. There is a basic trade-off between a high neutrino luminosity (which drives strong convection) and a low baryon loading, but convective instability will probably occur in a new-born neutron star (or torus) even when the neutrino flux is 2–3 orders of magnitude below its peak value, and the baryon ablation rate is suppressed accordingly. Finally, a strong ($B \sim 10^{16}$ G) small-scale magnetic field generated by the convective motions can trap most of the baryons near the stellar surface, thereby suppressing the steady mass-loss rate by a factor of 10 or more.

The rapid variability of some GRBs can, in principle, be accommodated in a cosmological fireball with a high bulk Lorentz factor (Paczyński 1990). We have nothing new to add to this question, except to note that variability of the emergent gamma-ray flux on very short time-scales is tied to a hydromagnetic instability occurring near the photosphere, *not* near the source. Although the spatial variation in the magnetic field that is needed to trigger the instability is generated near the source (either by rotation or by some more complicated process such as reconnection), the amplitude of the Alfvén turbulence must continually increase with radius in order to maintain a constant Compton y-parameter, which means that the turbulence must be generated locally. Highly intermittent reconnection will occur at neutral sheets whose mean separation exceeds the rest-frame horizon of the wind. Variability on time-scales much longer than the rotation period of the source is more likely caused by a variation in the ratio of Poynting flux to baryon mass flux. This could be most easily produced by a sudden increase in the spin-down torque due to a rearrangement of the external magnetic field of the neutron star or disc. A sudden change in the neutrino-driven mass ablation rate is less likely in the case of a neutron star undergoing secular cooling, but may occur in the case of a disc undergoing viscous spreading. We also note that a high baryon density has the effect of increasing the minimum variability time-scale at the Thomson photosphere. We predict a positive correlation between the variability of a burst and the position of the high-energy spectral cut-off (but not necessarily the break energy).

One potential inadequacy of the model is that it does not naturally produce the soft X-ray tails (and precursors) seen by *Ginga* (Murakami et al. 1992). In our opinion, these tails are the only piece of evidence that points directly to Galactic neutron stars as the source of the classical GRBs, since the colour temperatures and fluxes are reproduced (in a few cases) by radiation from a neutron star at the Eddington rate at a distance of 10 kpc (Murakami et al. 1992). A thermal photon spectrum will arise in an ultraluminous, magnetically dominated wind only if the level of Alfvén turbulence in the wind is low. Moreover, the photon temperature emerging from such a wind has a very strong dependence on γ_b for high baryon loading, $T(\text{observed}) \propto L_\gamma \propto \gamma_b^{10/3}$ (Section 5.5). So, while adiabatic softening from the temperature (25) to a temperature $T(\text{observed}) \sim 2$ keV (between a radius $\sim \pi\gamma_b^2 r_{lc}$ and the photosphere) could be reproduced with the baryon loading

$$\gamma_b = 0.20\, L_{50}^{-3/40} P_{-3}^{3/20} \left[\frac{T(\text{observed})}{2\ \text{keV}}\right]^{3/10} \gamma_{b,\text{ann}} = 19\, L_{50}^{1/8} P_{-3}^{-1/10} \left[\frac{T(\text{observed})}{2\ \text{keV}}\right]^{3/10}, \tag{78}$$

the range of allowed γ_b is rather small. This also suggests that lower frequency (e.g. optical) emission from a component of the wind with a very heavy mass loading is strongly suppressed by adiabatic losses. Most of the optical emission is presumably generated where the wind interacts with ambient material (e.g. Mészáros & Rees 1993). Notice that the soft X-ray *precursor* seen in a few bursts (Murakami et al. 1992) can be accommodated if the baryon mass-loss rate *decreases* with time, as is expected with ablation by neutrino cooling. The soft X-ray light curve is often observed to be much smoother than the light curve of the harder gamma-rays, suggesting that the harder component is beamed.

We close with the following question: can the same radiative mechanism be applied to gamma-ray bursts in the Galactic halo? One obvious drawback of a halo fireball model is that the luminosity of a burst is reduced by some nine orders of magnitude from that in a cosmological model, which means that the characteristic break energy (71) drops by a factor of ~200 to a value well below 1 MeV. This problem can be fixed by reducing the photon number flux by a comparable factor, which has the effect of raising $E_{\rm break}$ (equation 64). Another, more serious, drawback of a halo model is that the high-energy spectrum of the burst is cut off at a relatively low energy, since the bulk Lorentz factor inside the Thomson photosphere never exceeds $\gamma_{\rm b}^* \sim$ a few. This problem has been encountered, in a slightly different guise, in previous work on Galactic disc GRB sources (Vitello & Dermer 1991, and references therein). Without the advantage of relativistic bulk expansion, Comptonization of soft X-ray photons in neutron star magnetospheres requires large electron injection energies ($\gamma_{\rm e} \sim 10^3$) in order to produce power-law tails that extend well above 1 MeV. These models have difficulty in combining these tails with spectral breaks that lie in the vicinity of 1 MeV. We conclude that, if classical gamma-ray bursts are generated by mildly relativistic Alfvén turbulence, then extended high-energy spectra seem to require highly relativistic bulk expansion of the sort that is best achieved in cosmological sources.

ACKNOWLEDGMENTS

I thank Omer Blaes for many interesting discussions of cosmological gamma-ray bursts, which led to the work presented here, and Roger Blandford, Robert Duncan and Peter Goldreich for conversations. I also thank Michael Nowak and the referee, Martin Rees, for comments that helped to clarify the presentation. This work was funded by the NSERC of Canada.

REFERENCES

Arnett W. D., Bowers R. L., 1977, ApJS, 33, 415
Balbus S. A., Hawley J. F., 1991, ApJ, 376, 214
Band D. et al., 1992, in Paciesas W. S., Fishman G. J., Gamma Ray Bursts, Am. Inst. Phys., New York, p. 169
Band D. et al., 1993, ApJ, 413, 281
Bethe H. A., 1990, Rev. Mod. Phys., 62, 801
Bhattacharya D., van den Heuvel E. P. J., 1991, Phys. Rep., 203, 1
Blandford R. D., 1993, in Friedlander M., Gehrels N., Macomb D. J., eds, Compton Gamma-Ray Observatory Symposium. Am. Inst. Phys., New York, p. 992
Blandford R., Eichler D., 1987, Phys. Rep., 154, 1
Blandford R. D., Payne D. G., 1981, MNRAS, 194, 1033
Blandford R. D., Payne D. G., 1982, MNRAS, 199, 883
Blandford R. D., Znajek R. L., 1977, MNRAS, 179, 433
Brown G. E., Bethe H. A., 1994, ApJ, 423, 659
Burrows A., 1987, ApJ, 318, L57
Cattaneo F., Vainshtein S. I., 1991, ApJ, 376, L21
Cooper G., 1971, Phys. Rev., D3, 2312
Coroniti F. V., 1990, ApJ, 349, 538
Dar A., Kozlovsky B. Z., Nussinov S., Ramaty R., 1992, ApJ, 388, 164
Davies M. B., Benz W., Piran T., Thielemann F. K., 1993, preprint
De Jager O. C., Harding A. K., 1992, ApJ, 396, 161
Duncan R. C., Thompson C., 1992, ApJ, 392, L9 (DT92)
Duncan R. C., Shapiro S. L., Wasserman I., 1986, ApJ, 309, 141
Duncan R. C., Li H., Thompson C., 1993, in Friedlander M., Gehrels N., Macomb D. H., eds, Compton Gamma-Ray Observatory Symposium. Am. Inst. Phys., New York, p. 1074
Eichler D., 1993, ApJ, 419, 111
Fermi E., 1949, Phys. Rev., 75, 1169
Field G. B., Rogers R. D., 1993, ApJ, 403, 94
Fruchter A. S. et al., 1990, ApJ, 351, 642
Goldreich P., Julian W. H., 1969, ApJ, 157, 869
Goldreich P., Julian W. H., 1970, ApJ, 160, 971
Goldreich P., Reisenegger A., 1992, ApJ, 395, 250
Goodman J., 1986, ApJ, 308, L47
Goodman J., Xu G., 1993, Princeton preprint
Goodman J., Dar A., Nussinov S., 1987, ApJ, 314, L7
Haisch B. M., Strong K. T., Rodonò M., 1991, ARA&A, 29, 275
Hall D. S., 1989, Space Sci. Rev., 50, 219
Janka H.-T., 1991, A&A, 244, 378
Kennel C. F., Coroniti F. V., 1984a, ApJ, 283, 694
Kennel C. F., Coroniti F. V., 1984b, ApJ, 283, 710
Kompane'ets A. S., 1957, Sov. Phys. JETP, 4, 730
Kouveliotou C. et al., 1994, ApJ, 422, L59
Kuznetsov A. V., Sunyaev R. A., Terekhov O. V., Yakubtsev L. A., Barat C. et al., 1986, Sov. Astron. J. Lett., 12, 311
Li Z. Y., Chiueh T., Begelman M. C., 1992, ApJ, 394, 459
Lightman A. P., 1981, ApJ, 244, 392
Mao S., Paczyński B., 1992, ApJ, 389, L13
Mayle R. W., Wilson J. R., 1989, preprint
Meegan C. A. et al., 1992, Nat, 355, 143
Mészáros P., Rees M. J., 1992, MNRAS, 257, 29P
Mészáros P., Rees M. J., 1993, ApJ, 418, L59
Miller D. S., Wilson J. R., Mayle R. W., 1993, ApJ, 415, 278
Murakami T. et al., 1992, in Paciesas W. S., Fishman G. J., Gamma Ray Bursts. Am. Inst. Phys., New York, p. 169
Mushotzky R. F., Done C., Pounds K. A., 1993, ARA&A, 31, 717
Narayan R., Popham R., 1989, ApJ, 346, L25
Narayan R., Paczyński B., Piran T., 1992, ApJ, 395, L83
Paczyński B., 1986, ApJ, 308, L43
Paczyński B., 1990, ApJ, 363, 218
Paczyński B., Rhoads J. E., 1993, ApJ, 418, L5
Pallavicini R., Gollub L., Rosner R., Vaiana G. S., Ayres T. R., Linsky J. L., 1981, ApJ, 248, 279
Peebles P. J. E., 1971, Physical Cosmology. Princeton Univ. Press, Princeton, NJ
Petschek H. E., 1964, in Hess W. N., ed., AAS–NASA Symp., The Physics of Solar Flares. NASA SP-50, p. 425
Pozdnyakov L. A., Sobol' I. M., Sunyaev R. A., 1983, Astrophys. Space Phys. Rev., 2, 189
Priest E. R., Forbes T. G., 1992, J. Geophys. Res. Space Phys., 97, 16757
Rasio F. A., Shapiro S. L., 1992, ApJ, 401, 226
Rasio F. A., Shapiro S. L., 1994, preprint
Rees M. J., 1984, ARA&A, 22, 471
Rees M. J., Gunn J. R., 1974, MNRAS, 167, 1
Romanova M. M., Lovelace R. V. E., 1992, A&A, 262, 26
Sang Y., Chanmugam G., 1987, ApJ, 323, L61
Semenov V. S., Bernikov L. V., 1991, Sov. Phys. JETP, 72, 813
Shaefer B. E. et al., 1993, ApJS, 92, 285
Shapiro S. L., Teukolsky S. A., 1983, Black Holes, White Dwarfs, and Neutron Stars. John Wiley, New York, p. 361
Shapiro S. L., Lightman A. P., Eardley D. M., 1976, ApJ, 204, 187
Shemi A., Piran T., 1990, ApJ, 365, L55
Simon T., 1990, ApJ, 359, L51
Simon T., Herbig G., Boesgaard A. M., 1984, ApJ, 293, 551
Sommer M. et al., 1994, ApJ, 422, L63
Stern R. A., Haisch B. M., Uchida Y., Walter F., Veale A., Brown A., Vilhu O., Hannikainen D., 1991, in Wallerstein G., ed., ASP Conf. Ser. Vol. 9, Cool Stars, Stellar Systems and the Sun. Astron. Soc. Pac., San Francisco, p. 227
Sunyaev R. A., Titarchuk L. G., 1980, A&A, 86, 121

Sunyaev R. A., Zel'dovich Ya. B., 1980, ARA&A, 18, 537
Svensson R., 1987, MNRAS, 227, 403
Thompson C., Duncan R. C., 1993, ApJ, 408, 194 (TD93)
Thompson C., Duncan R. C., 1994, in Backer D., Fruchter A. S., Tavani M., eds, Proc. Workshop on Millisecond Pulsars, Aspen, Colorado. Astron. Soc. Pac., San Francisco
Thompson C., Blandford R. D., Evans C. R., Phinney E. S., 1994, ApJ, 422, 304
Tout C. A., Pringle J. E., 1992, MNRAS, 256, 269
Usov V. V., 1992, Nat, 357, 472
Usov V. V., 1994, MNRAS, 267, 1035
Vitello P., Dermer C. D., 1991, ApJ, 374, 668
Weber E. J., Davis L., Jr, 1967, ApJ, 148, 217
Wijers R. A. M., Verbunt F., Bhattacharya D., Hartman J. W., 1993, in Van Riper K. A., Epstein R., Ho C., eds, Isolated Pulsars. Cambridge Univ. Press, Cambridge, p. 83
Wilson D. B., Rees M. J., 1978, MNRAS, 185, 297
Wilson J. R., 1985, in Centrella J. M., LeBlanc J. M., Bowers R. L., eds, Numerical Astrophysics. Jones & Bartlett, Boston, p. 422
Witti J., Janka H.-Th., Takahashi K., 1994, A&A, 286, 841
Woosley S. E., 1993, ApJ, 405, 273