



**HAL**  
open science

## Connecting steady emission and very high energy flaring states in blazars: the case of Mrk 421

A. Dmytriiev, H. Sol, A. Zech

► **To cite this version:**

A. Dmytriiev, H. Sol, A. Zech. Connecting steady emission and very high energy flaring states in blazars: the case of Mrk 421. *Monthly Notices of the Royal Astronomical Society*, 2021, 505 (2), pp.2712-2730. 10.1093/mnras/stab1445 . hal-03261180

**HAL Id: hal-03261180**

**<https://hal.science/hal-03261180>**

Submitted on 3 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Connecting steady emission and very high energy flaring states in blazars: the case of Mrk 421

A. Dmytriiev \*, H. Sol\* and A. Zech\*

LUTH, Observatoire de Paris, Université de Paris, CNRS, PSL University, 5 Pl. Jules Janssen, F-92195 Meudon, France

Accepted 2021 May 11. Received 2021 March 29; in original form 2020 October 19

## ABSTRACT

Various attempts have been made in the literature at describing the origin and the physical mechanisms behind flaring events in blazars with radiative emission models, but detailed properties of multiwavelength (MWL) light curves still remain difficult to reproduce. We have developed a versatile radiative code, based on a time-dependent treatment of particle acceleration, escape, and radiative cooling, allowing us to test different scenarios to connect the continuous low-state emission self-consistently with that during flaring states. We consider flares as weak perturbations of the quiescent state and apply this description to the 2010 February MWL flare of Mrk 421, the brightest very high energy (VHE) flare ever detected from this archetypal blazar, focusing on interpretations with a minimum number of free parameters. A general criterion is obtained, which disfavors a one-zone model connecting low and high state under our assumptions. A two-zone model combining physically connected acceleration and emission regions yields a satisfactory interpretation of the available time-dependent MWL light curves and spectra of Mrk 421, although certain details remain difficult to reproduce. The two-zone scenario finally proposed for the complex quiescent and flaring VHE emitting region involves both Fermi-I and Fermi-II acceleration mechanisms, respectively, at the origin of the quiescent and flaring emission.

**Key words:** acceleration of particles – radiation mechanisms: non-thermal – BL Lacertae objects: individual: Mrk 421.

## 1 INTRODUCTION

### 1.1 Mrk 421 at VHE $\gamma$ -rays

Markarian 421 (Mrk 421, RA = 11<sup>h</sup> 04<sup>m</sup> 27<sup>s</sup>, Dec = +38° 12′ 32″,  $z = 0.031$ ) is generally the brightest extragalactic  $\gamma$ -ray source in the very-high-energy (VHE; energy range above about 100 GeV)  $\gamma$ -ray sky. It is also the most nearby representative of the blazar type of active galactic nuclei (AGNs), the sub-class of radio-loud AGN with jets aligned along the line of sight, and one of the best-studied high-frequency-peaked BL Lac objects (HBLs). The flux from this type of sources is highly variable, with flux levels typically changing by more than one order of magnitude during flaring activity.

To reveal the origin of the flaring behaviour of blazars, one needs to understand the changes of the physical conditions in the source between the quiescent and flaring states. This requires that timing and spectral properties of the source are well constrained from observations in these two states. In the VHE  $\gamma$ -ray band, where the fastest variability is observed, this is generally not the case. Imaging Atmospheric Cherenkov Telescopes (IACTs), which are the most sensitive observational tools in this energy range, have fields of view of a few degrees, so the continuous long-term monitoring of a large number of sources distributed across the sky is not possible. Despite the fast development of target of opportunity (ToO) observations, it remains

difficult to promptly detect and observe flaring events, and the vast majority of the outbursts observed in VHE  $\gamma$ -rays are incomplete in time coverage. Moreover, MWL alerts and campaigns request the participation of many different ground-based and space instruments, which are complex to coordinate. In this context, Mrk 421 is one of the most observed sources at VHE, providing particularly rich data sets.

The long-term variability of Mrk 421 in the VHE band was studied first with the Whipple telescope over a 14-yr period (Acciari et al. 2014). Significant flux variations were found on time-scales ranging from a few minutes to years. Evidence was seen for correlations with the X-ray flux at monthly and yearly time-scales. A study of the source’s long-term MWL behaviour from 2007 to 2009 including data from the MAGIC IACTs found VHE flux variations in the range from  $1.3 \times 10^{-11} < F(>400 \text{ GeV}) < 3.1 \times 10^{-10} \text{ ph cm}^{-2} \text{ s}^{-1}$  over this period (Ahnen et al. 2016). During both low and high states the observed variability is higher in the high-energy bands than in the low-energy bands. A direct correlation of the VHE emission with the X-ray band was confirmed, while the authors found no significant correlation with the radio and optical bands.

Several flaring events from this source were studied for instance by Cao & Wang (2013) (June 2008 flare), Aleksić et al. (2015) (March 2010 flare), Hovatta et al. (2015) (2012 and 2013 flares), MAGIC collaboration (2020) (2013 April flare). To explain the observed correlated behaviour between X-rays and  $\gamma$ -rays during flux fluctuations, a synchrotron self-Compton (SSC) emission scenario is usually preferred.

\* E-mail: [anton.dmytriiev@obspm.fr](mailto:anton.dmytriiev@obspm.fr) (AD); [helene.sol@obspm.fr](mailto:helene.sol@obspm.fr) (HS); [andreas.zech@obspm.fr](mailto:andreas.zech@obspm.fr) (AZ)

## 1.2 2010 February VHE flare of Mrk 421

We focus in this paper on the brightest VHE flare as of yet, reaching a level of about 27 Crab units above 1 TeV, observed in 2010 February and followed with a range of instruments from the radio band up to VHE  $\gamma$ -rays as described by Tluczykont (2011), Shukla et al. (2012), Singh et al. (2015), and Abeysekera et al. (2020). The flux of the source was seen to vary over seven days from the optical to the VHE band, with a peak on MJD 55243 or MJD 55244 and a possible time delay of the low-energy signal relatively to the VHE  $\gamma$ -rays. The X-ray flux increased by more than a factor of two in one day and spectral hardening with increasing flux was observed in X-rays and  $\gamma$ -rays. The detailed analysis of MWL variability and correlation studies recently published by the VERITAS and MAGIC collaboration with their MWL partners (Abeysekera et al. 2020) reports short time-scale variability with the emission rapidly varying during the main flare on the 10 min time-scale, and a complex VHE versus X-ray flux relationship. We will focus here on the MWL behaviour of the flare emission at a daily time-scale and do not try to directly reproduce very rapid variability.

## 1.3 Physical modelling of blazar flares

To explain the emergence of flares within relativistic jets, one generally distinguishes two possible types of scenarios. In the first type, variations of macrophysical properties of the emitting region, like its global geometry and kinematics, are responsible for launching outbursts. For instance, the relativistic Doppler factor can be increased by a change of viewing angle or bulk Lorentz factor, which can lead to stronger emission boosting and launch flares (e.g. Casadio et al. 2015; Larionov et al. 2016; Raiteri et al. 2017). This happens if the emitting region is moving along a curved trajectory, due to jet bending (Abdo et al. 2010) or its helical configuration (Villata & Raiteri 1999). In the second type, the observed flux variability is considered to be due to the microphysics inside the VHE emitting zone and the subsequent evolution of the population of radiating particles caused by various physical processes, for example enhanced injection (e.g. Mastichiadis & Kirk 1997), and different particle acceleration mechanisms due to the development of shocks (e.g. Marscher & Gear 1985; Sikora et al. 2001; Böttcher & Baring 2019), turbulence (e.g. Boutelier, Henri & Petrucci 2008; Tammi & Duffy 2009; Tramacere, Massaro & Taylor 2011; Shukla et al. 2016) or magnetic reconnection (e.g. Giannios, Uzdensky & Begelman 2009; de Gouveia Dal Pino et al. 2010; Shukla et al. 2018; Shukla & Mannheim 2020). Emission models often combine the two types of mechanisms in order to reproduce extreme flares (e.g. Katarzyński, Sol & Kus 2001; Marscher 2014).

## 1.4 Previous interpretations of the 2010 February flare

A first interpretation of the remarkable 2010 flare of Mrk 421 using spectral modelling is given by Shukla et al. (2012) and Singh et al. (2015, 2017). Singh et al. (2015) model the daily averaged spectral energy distribution (SED) on MJD 55243 with a standard one-zone SSC scenario, assuming an instantaneous electron distribution that follows a broken power law. The variability time-scale is used to constrain the source parameters relying on the usual light traveltime argument.

A similar approach by Shukla et al. (2012) provides instantaneous models for several SED data sets before, during, and after the flare. The authors distinguish between a peak in X-rays and soft  $\gamma$ -rays on MJD 55243 and a peak possibly delayed in the VHE emission on

MJD 55244. In their scenario, the flaring state arises from a change in several parameters, including Doppler factor, magnetic field strength, electron energy density, and spectral index, which is attributed to electron acceleration in a strong shock.

Singh et al. (2017) propose a first model of the MWL light curve during the flare, which they find to be asymmetric in the high-energy bands, with a rise that is faster than the decay. They model the flux evolution detected by *Swift*-XRT, MAXI, *Fermi*-LAT, and TACTIC with a one-zone model by adjusting a time-dependent injection function, assuming a constant spectral index of about 2.0.

The most recent interpretation of the MWL spectral evolution during the flare is given by Banerjee et al. (2019), based on a time-dependent internal shock model (Joshi & Böttcher 2011). Four spectral states before and during the flare were fit individually by varying key parameters of this multizone model. The authors find notably that the particle injection spectrum hardens during the flare, which they interpret as a shift from a dominant shock acceleration mechanism during the low state to stochastic acceleration during the flaring state.

## 1.5 The present model for the 2010 February flare

In this paper, we explore the approach based on the evolution of the particle distribution, assuming constant physical parameters of the emitting zone (such as its size, Doppler factor, and magnetic field strength), and propose time-dependent SSC scenarios trying to connect self-consistently the long-term low-state emission from Mrk 421 as described by Abdo et al. (2011), with the prominent 2010 February flare. To understand the nature of the outburst and the relevant physical processes involved, we perform detailed physical modelling of the data set, including stochastic and shock acceleration in addition to mere particle injection and produce fits of the MWL light curves in optical to VHE bands as well as of spectral measurements during different activity states, having in mind to propose a coherent global picture of the source with a physical model in which the variability pattern observed during the flare arises naturally from the quiescent state of the source due to a weak perturbation in or around the long-term emission region.

The data sets on Mrk 421 studied in this work are presented in Section 2. In Section 3, we introduce the time-dependent emission code we have developed, which serves as a building block for the scenarios discussed in the following sections. To limit the number of free parameters, homogeneous one-zone models are explored first, and then extended to two zones when necessary to provide a satisfactory description of the data set. In Section 4, we analyse and reproduce the quiescent state of Mrk 421. A criterion to test the validity of one-zone models is established in Section 5 for individual SSC flares. Section 6 shows how a two-zone model can reproduce the MWL light curves of the flare. A general discussion of the results and perspectives is presented in Section 7.

## 2 OBSERVATIONAL DATA

### 2.1 Data set for the low state

Abdo et al. (2011) present a composite MWL data set of Mrk 421 in a low activity state, obtained as a result of a 4.5-month-long observational campaign on the source (2009 January 19–June 1). Many instruments were participating in the campaign, including the VLBA, radio telescopes from the F-GAMMA program, optical and near-infrared telescopes from the GASP-WEBT program, *Swift*, RXTE, *Fermi*-LAT, and MAGIC, among others. During this time

period, the source was found in a relatively low flux state, and displayed almost no variability in all energy bands. The combined time-averaged measurement reported by the authors thus serves as a good proxy for a long-term quiescent or steady-state SED of Mrk 421.

## 2.2 Data set for the 2010 February flare

A giant flare was observed during the period 2010 February 10–23 (MJD 55237–55250). The source showed variability from the optical to the VHE  $\gamma$ -ray range. In X-rays and in the VHE band, the energy flux increased by a factor of about 5 to 10, reaching a peak around 2010 February 16 (MJD 55243.5–55244.0). The X-ray flux was found to correlate with the one in the TeV band (Shukla et al. 2012). Variability was observed at two time-scales:  $\sim 1$  d (in all energy bands), and  $\sim 1$  h (intranight variability) at TeV energies (Shukla et al. 2012; Abeysekara et al. 2020). A secondary flaring event is observed after the main flare in X-rays at MJD 55249, however it is not significant at MeV–GeV energies in the *Fermi* light curve. The flux increase in the optical *V*-band is very modest at around 20–30 per cent and the variability in the radio band is negligible (Shukla et al. 2012).

We have compiled the published light curves over all wavelength bands, as well as the available spectra. VHE observations of the outburst were performed with H.E.S.S. from MJD 55245.0 to MJD 55247.0 (Tluczykont 2011), and VERITAS around MJD 55244.3 (Fortson, VERITAS Collaboration & Fermi-LAT Collaborators 2012; Abeysekara et al. 2020). VERITAS also monitored the source during the three following nights. Unfortunately, these instruments observed neither the rise of the flare nor the very peak, with VERITAS starting data-taking roughly one day after the estimated flux maximum, and H.E.S.S. – around 1.5 d after the estimated peak. More complete time coverage in the VHE range, albeit with smaller sensitivity, was achieved by the HAGAR array (Shukla et al. 2012), monitoring the source during February 13–19, and by the TACTIC Cherenkov telescope (Singh et al. 2015), observing during February 10–23. The VHE flux recorded with the TACTIC telescope appeared significantly lower than the flux observed by the VERITAS and H.E.S.S. telescope arrays when compared in the same energy range and during the same time period, assuming the average spectral shape measured with H.E.S.S. We ascribe this discrepancy to uncertainties in the absolute calibration and rescale the TACTIC light curve by a constant factor of 5.7 to ensure consistency with VERITAS and H.E.S.S. fluxes.

Data during the flare from the *Fermi*-LAT instrument in the MeV–GeV  $\gamma$ -ray band were published by Singh et al. (2015) (who used instrument response functions P7SOURCE\_V6 in their analysis) and by Abeysekara et al. (2020) (who used more recent instrument response functions P8R2\_SOURCE\_V6), X-ray data from *Swift*-XRT by Shukla et al. (2012) and Singh et al. (2015), *Swift*-BAT data by Shukla et al. (2012), MAXI data by Singh et al. (2015), RXTE-PCA and RXTE-ASM data by Shukla et al. (2012). Optical data are available from the SPOL telescope and radio data from OVRO and were taken from Shukla et al. (2012).

To constrain the time-dependent SED, we compare our model results to spectra measured with XRT in soft X-ray and *Swift*-BAT in the hard X-ray band near the peak of the flare (2010 February 16), to the *Fermi*-LAT uncertainty band for the spectrum (all taken from Singh et al. 2015), as well as to spectra from H.E.S.S. (time-averaged SED for the period 2010 February 17–20) (Tluczykont 2011) and VERITAS (based on 5 h of data taken on 2010 February 17, roughly one day after the estimated peak of the flare) (Fortson et al. 2012).

To compare the published count rates from the RXTE–ASM and *Swift*–BAT instruments to the light curves from the other wavelength bands, they were converted into energy fluxes first. The RXTE–ASM count rates were converted following the prescriptions by Grimm, Gilfanov & Sunyaev (2002) and Chitnis et al. (2009) for a Crab-like spectrum. The difference between the Crab spectrum and an average spectrum of Mrk 421 in this energy range was evaluated to add a systematic error of about 2 per cent to the uncertainty of the energy flux. It was also verified that photoabsorption on the Galactic hydrogen column is negligible in this energy range.

For the conversion of the *Swift*–BAT count rates to a photon flux, the count rate at the peak is normalized with the spectrum presented by Singh et al. (2015), which we fit with a power law and integrate over the energy range of 15 to 50 keV. A systematic uncertainty from the fitting procedure is added to the error on the resulting photon flux.

## 3 THE EMBLEM TIME-DEPENDENT BLAZAR MODEL

We adopt the conventional leptonic SSC scenario for the origin of blazar  $\gamma$ -ray emission, assuming it originates from a spherical region of radius  $R_b$  (a ‘blob’), filled with a homogeneous relativistic electron–positron plasma embedded in a tangled magnetic field  $B$  with a uniform flux density. The plasma blob is relativistically moving along the jet axis with a Doppler factor  $\delta_b$ . From now on we will refer to electrons and positrons as simply electrons. The electron population in the emitting zone is evolving due to several processes. Particles are injected into the blob with a spectrum  $Q_{\text{inj}}(\gamma)$  which may depend on time, and may gain energy due to acceleration by shock (Fermi-I) or stochastic (Fermi-II) processes. The electrons confined in the blob radiate synchrotron and inverse Compton (IC) emission, and cool through radiative losses, comprising synchrotron and inverse Compton cooling. We also take into account synchrotron self-absorption. The particles escape the emitting zone at a characteristic time-scale  $t_{\text{esc}}$ . We assume that the size of the blob remains constant at first order, which allows us to drastically reduce the number of free parameters. We also neglect adiabatic losses assuming that during the quiescent state the emitting zone is confined inside the jet (e.g. by its pressure external to the blob) with constant size, and internal  $\gamma$ – $\gamma$  absorption, which is typically negligible in HBLs (e.g. Katarzyński et al. 2001). Also, we disregard the emission from the extended jet, which we assume is only significant at very low energies. We treat the flaring behaviour as originating from the varying spectrum of the electron population in the blob.

The time evolution of the electron spectrum  $N_e(\gamma, t)$  in the emitting zone is governed by a kinetic (Fokker–Planck) equation representing a continuity equation in phase space. Its general form, taking into account the above-mentioned physical processes in the blob, for the case of ‘hard-sphere’ turbulence (see sub-section 3.3) is (e.g. Kardashev 1962; Tramacere et al. 2011):

$$\begin{aligned} \frac{\partial N_e(\gamma, t)}{\partial t} = & \frac{\partial}{\partial \gamma} [(b_c \gamma^2 - a\gamma - 2D_0\gamma) \cdot N_e(\gamma, t)] \\ & + \frac{\partial}{\partial \gamma} \left( D_0 \gamma^2 \frac{\partial N_e(\gamma, t)}{\partial \gamma} \right) - \frac{N_e(\gamma, t)}{t_{\text{esc}}} + Q_{\text{inj}}(\gamma, t). \end{aligned} \quad (1)$$

The physical processes behind the different terms on the right-hand side are described in the following subsections.

### 3.1 Electron cooling

The term  $b_c \gamma^2$  corresponds to the total radiative cooling rate, comprising the synchrotron and the inverse Compton cooling rate:

$$b_c \gamma^2 = |\dot{\gamma}_{\text{syn}}| + |\dot{\gamma}_{\text{IC}}|$$

The synchrotron cooling rate is (e.g. Chiaberge & Ghisellini 1999):

$$|\dot{\gamma}_{\text{syn}}| = \frac{4\sigma_T}{3m_e c} \gamma^2 U_B,$$

where  $U_B = \frac{B^2}{2\mu_0}$  is the magnetic energy density,  $\sigma_T$  is the Thomson cross-section,  $m_e$  is the electron rest mass,  $c$  is the vacuum speed of light.

The inverse Compton cooling rate is given by Moderski et al. (2005):

$$|\dot{\gamma}_{\text{IC}}| = \frac{4\sigma_T}{3m_e c} \gamma^2 \int_{\epsilon'_{\text{min}}}^{\epsilon'_{\text{max}}} f_{\text{KN}}(4\gamma\epsilon') u'_{\text{syn}}(\epsilon') d\epsilon',$$

where  $\epsilon' = \frac{h\nu'}{m_e c^2}$  is the energy of seed (synchrotron) photons in the reference frame of the emitting zone in the units of the electron rest energy,  $u'_{\text{syn}}(\epsilon')$  is the distribution of the energy density of the synchrotron photons, and the function  $f_{\text{KN}}(x)$  is approximated as (including Klein–Nishina effects):

$$f_{\text{KN}}(x) \simeq \begin{cases} (1+x)^{-1.5}, & \text{for } x < 10^4 \\ \frac{9}{2x^2} (\ln x - 11/6), & \text{for } x \geq 10^4 \end{cases}$$

The term  $b_c$  in the total cooling rate is thus

$$b_c = \frac{4\sigma_T}{3m_e c} \left[ U_B + \int_{\epsilon'_{\text{min}}}^{\epsilon'_{\text{max}}} f_{\text{KN}}(4\gamma\epsilon') u'_{\text{syn}}(\epsilon') d\epsilon' \right] \quad (2)$$

and the characteristic cooling time of electrons is  $t_{\text{cool}}(\gamma) = \frac{1}{b_c \gamma}$ .

It is worth to note, that the standard kinetic equation (equation 1) could be formally not applicable in the case where the inverse Compton cooling in Klein–Nishina regime becomes important, since the equation (1) is derived assuming that the electrons lose only a small fraction of their energy in one interaction, while this is no longer the case in Klein–Nishina regime. However, inverse Compton cooling is usually negligible in HBLs, so that the standard kinetic equation can be applied to Mrk 421.

### 3.2 Shock acceleration

Fermi-I (or diffusive shock) acceleration operates at the fronts of hydrodynamical shock waves, in the presence of velocity discontinuities. The principal mechanism of acceleration of relativistic charged particles by a strong shock is discussed by Bell (1978). As a collision-less shock propagates through magnetized plasma, relativistic particles scatter on some turbulences or Alfvén waves (Wentzel 1974) in the downstream and upstream regions, gradually gaining energy at each crossing of the shock front. The same scattering processes which entrap the particles near the shock are also responsible for their escape. The downstream plasma is receding from the shock front, leading to advection of the accelerated particles away from the shock, since their velocity distribution is isotropic in the medium frame. The spectrum of the particles escaping the shock follows a power law  $dN_e/d\gamma \propto \gamma^{-\alpha_{\text{es}}}$  with an index  $2 \leq \alpha_{\text{es}} \leq 2.5$ , depending on the gas compression ratio.

Shock acceleration is a ubiquitous phenomenon thought to occur in many astrophysical systems, in particular in AGN jets (e.g. Marscher & Gear 1985). It is therefore natural to consider the possibility that

the blazar outbursts could be triggered by a shock passing through the emitting zone. In addition, continuously operating on a long-term basis, the Fermi-I process could also serve as an efficient pre-acceleration mechanism supplying high-energy particles to the blob. Pre-accelerated particles can be further re-accelerated by another shock or by second-order Fermi mechanisms.

We treat the shock acceleration process with the kinetic approach of equation (1). The Fermi-I process is considered as a systematic energy gain, and is described in the kinetic equation by the term  $a\gamma$ , which is the Fermi-I acceleration rate, a quantity proportional to the particle energy gain per unit of time:  $\dot{\gamma}_{\text{FI}} = a\gamma$ . The characteristic time-scale of the shock acceleration is  $t_{\text{FI}} = 1/a$ . Throughout the paper, the  $a$ -term is put equal to zero when the shock acceleration is not active, and can be suddenly activated as needed to start the acceleration process. This introduces an additional important parameter  $t_{\text{dur,FI}}$  for the duration of the acceleration phase, namely  $t_{\text{dur,FI}} = t_{\text{cs}}$  for the crossing time or the lifetime of the shock defined in the source frame.

### 3.3 Turbulence and stochastic acceleration

The presence of turbulence in the region can also lead to Fermi-II acceleration of particles. The energy is injected into the region at the largest spatial (stirring) scale, comparable to the size of the region, and cascades down to smaller scales, until the minimal one at which viscosity losses become dominant. The turbulent motion in the magnetized plasma produces a stochastic component of the magnetic field  $\delta B$  in addition to the main (ordered) component  $B_0$ , which perturbs the plasma and excites Alfvén waves. In the quasi-linear framework that we consider here (e.g. Schlickeiser 1989; Jaekel & Schlickeiser 1992), the magnetohydrodynamic (MHD) turbulence is described by a combination of Alfvén waves with different wave numbers, forming a continuous wave spectrum. Particles of the plasma interact with the Alfvén waves and may exchange energy and momentum, leading to a gradual energy gain in a stochastic manner (see e.g. Dermer, Miller & Li 1996). The momentum diffusion coefficient describing momentum–energy gain by a particle is controlled by the wave-turbulence power spectrum, which has a form  $W(k) \propto k^{-q}$ , where  $k = 2\pi/\lambda$  is the wavenumber. The power spectrum is normalized as follows,  $\int_{k_{\text{min}}}^{k_{\text{max}}} W(k) dk = \frac{\delta B^2}{2\mu_0}$  which is the total energy density stored in the magnetic fluctuations. The minimum and the maximum wavenumbers correspond to the longest ( $\lambda_{\text{max}}$ ) and the shortest wavelength ( $\lambda_{\text{min}}$ ) in the Alfvén spectrum accordingly. The spectral index  $q = 3/2$  for the Kraichnan turbulence,  $q = 5/3$  for the Kolmogorov turbulence, and  $q = 2$  for the ‘hard-sphere’ approximation we adopt here to describe the Fermi-II acceleration (see e.g. Zhou & Matthaeus 1990; Asano & Hayashida 2018). The ‘hard-sphere’ turbulent spectrum is chosen as it favours the production of the brightest flares due to the most efficient re-acceleration of high-energy particles (e.g. Becker, Le & Dermer 2006). The choice is additionally justified based on the observational properties of the Mrk 421 2010 February flare (cf. sub-section 6.2).

The momentum diffusion coefficient for the process of stochastic acceleration in quasi-linear theory is given by (e.g. Schlickeiser 1989; O’Sullivan, Reville & Taylor 2009):

$$D_{p,\text{FI}}(p) \approx \beta_A^2 \frac{\delta B^2}{B_0^2} \left( \frac{r_L}{\lambda_{\text{max}}} \right)^{q-1} \frac{cp^2}{r_L}, \quad (3)$$

where  $\beta_A$  is the Alfvén speed in the units of the speed of light,  $r_L = \frac{pc}{eB_0}$  is the Larmor radius, and  $p = \sqrt{\gamma^2 - 1} m_e c$  is the electron momentum. The quantity  $\frac{\delta B^2}{B_0^2}$  is commonly referred to as ‘turbulence

level'. The characteristic time-scale of Fermi-II acceleration process is then

$$t_{\text{FII}} = \frac{p^2}{D_{p,\text{FII}}(p)} = \frac{1}{\beta_A^2} \frac{B_0^2}{\delta B^2} \frac{\lambda_{\text{max}}}{c} \left( \frac{r_L}{\lambda_{\text{max}}} \right)^{2-q}. \quad (4)$$

It scales with the particle momentum–energy as  $t_{\text{FII}} \propto p^{2-q}$ . The quasi-linear approach of Schlickeiser (1989) provides rather accurate (with an order of magnitude precision compared to numerical test-particle simulations) description of stochastic particle acceleration in the case of non-relativistic Alfvén speeds  $\beta_A \ll 1$  and low turbulence levels  $\delta B \ll B_0$ , and for mildly relativistic Alfvén waves and turbulence levels comparable to unity (O’Sullivan et al. 2009).

In the case of ‘hard-sphere’ turbulence, the stochastic acceleration time-scale  $t_{\text{FII}}$  is energy-independent, with

$$D_{p,\text{FII}}(p) \approx \beta_A^2 \left( \frac{\delta B}{B_0} \right)^2 \left( \frac{\lambda_{\text{max}}}{c} \right)^{-1} p^2 \equiv D_0 p^2$$

and

$$t_{\text{FII}} = \frac{1}{\beta_A^2} \left( \frac{B_0}{\delta B} \right)^2 \frac{\lambda_{\text{max}}}{c} = \frac{1}{D_0}. \quad (5)$$

As one can see, the Fermi-II time-scale is controlled by the turbulence level. In case it varies with time due to the evolution of the turbulence, the Fermi-II time-scale also changes with time. To estimate  $\beta_A$ , the energy density can be evaluated directly from the electron spectrum, assuming that the relativistic electrons dominate the total energy density  $\varepsilon$  and that for ultrarelativistic particles the pressure  $P = \frac{1}{3}\varepsilon$ . With these assumptions, the Alfvén speed for the case of relativistic MHD is given by (Gedalin 1993)

$$\beta_A = \frac{1}{\sqrt{1 + \frac{4\mu_0\varepsilon}{3B_0^2}}}. \quad (6)$$

The process of stochastic acceleration of electrons in the case of ‘hard-sphere’ turbulence is described in the kinetic equation (equation 1) by two terms, the first one  $-\frac{\partial}{\partial\gamma}(2D_0\gamma N_e(\gamma, t))$  is due to drift of electrons to higher Lorentz factors, with  $2D_0\gamma$  being proportional to particle energy gain per unit of time, and the second one  $\frac{\partial}{\partial\gamma}(D_0\gamma^2 \frac{\partial N_e(\gamma, t)}{\partial\gamma})$  describes the diffusion of the electron distribution in Lorentz factor space, where  $D_0\gamma^2$  is the energy diffusion coefficient.

Just like the Fermi-I process, the Fermi-II process in our model can be activated and deactivated when necessary. When the acceleration by turbulence is not active, the  $D_0$ -term is set to zero. This introduces again an additional free parameter  $t_{\text{dur,FII}}$ , which is the duration of the turbulent acceleration phase. In sub-section 6.2 this parameter is further decomposed into two parameters  $t_{\text{turb,r}}$  and  $t_{\text{turb,d}}$ , the rise and decay times of the turbulence.

Fermi-II acceleration mechanism is a universal process expected to operate in relativistic jets. In particular, it could work inside or in the vicinity of the emitting zone. Spontaneous formation of turbulence in the magnetized plasma of the blob or surrounding it results in turbulent magnetic fields. An attractive feature of the Fermi-II process, is that it is capable of producing electron spectra  $dN_e/d\gamma \propto \gamma^{-\alpha_{\text{es}}}$  with an index much harder than  $\alpha_{\text{es}} = 2$  (Virtanen & Vainio 2005). Also, the turbulent acceleration can serve as an efficient mechanism for re-acceleration of pre-accelerated particles.

### 3.4 Escape and injection of particles

The term  $\frac{N_e(\gamma, t)}{t_{\text{esc}}}$  in the equation (1) describes the escape of particles from the blob, with a characteristic escape time-scale  $t_{\text{esc}}$ , which in

general case depends on the particle energy. In the case of negligible or weak turbulence in the emission zone, particles escape it freely, and the escape time-scale is  $t_{\text{esc}} \sim 1R_b/c$ . Particles that undergo stochastic acceleration escape the region at a longer time-scale due to diffusion process. The spatial diffusion coefficient and the momentum diffusion coefficient are linked as  $D_x D_p \approx \beta_A^2 p^2$  (Skilling 1975). The escape time-scale from a turbulent region is therefore related to the stochastic acceleration time-scale as (Tramacere et al. 2011)

$$t_{\text{esc}}^{(\text{turb})} = \frac{R_t^2}{c^2 \beta_A^2 t_{\text{FII}}}, \quad (7)$$

where  $R_t$  is the size of the turbulent zone. For ‘hard-sphere’ turbulence,  $t_{\text{esc}}$  becomes energy-independent:

$$t_{\text{esc}}^{(\text{turb})} = \left( \frac{R_t}{c} \right)^2 \left( \frac{\delta B}{B_0} \right)^2 \frac{c}{\lambda_{\text{max}}}. \quad (8)$$

Finally the term  $Q_{\text{inj}}(\gamma, t)$  is the number of particles injected in a unit volume per unit time and per unit of Lorentz factor interval. It can be implemented as time-dependent injection spectrum. Particles may be injected continuously over a given time interval, or in an impulsive manner.

### 3.5 Numerical implementation

Following the kinetic approach discussed above, we developed a numerical code ‘EMBLEM’ (Evolutionary Modelling of BLOB Emission) for time-dependent modelling of blazar emission during flares. We use a fully implicit difference scheme by Chang & Cooper (1970) to numerically solve the kinetic equation (equation 1) and retrieve the time evolution of the electron spectrum on a time-grid. For the case without any Fermi-II acceleration term, we use a particular case of the Chang and Cooper scheme described in Chiaberge & Ghisellini (1999). We impose boundary conditions ensuring particle conservation in case of no injection or escape, following Park & Petrosian (1996). These boundary conditions prohibit leak of particles through the lower and upper boundaries of the Lorentz factor space that might occur due to processes merely changing particle energy (acceleration and cooling). As a result, the number of particles in the system is determined only by the injection and escape processes, while the acceleration and cooling processes cause particles to migrate between energy bins and cannot push particles beyond the user-defined domain of Lorentz factors. Throughout the paper, we define this domain wide enough in order to not artificially restrict the particle energy gain or loss processes. The SED of the emission from the blob is calculated at each time-step. For this, the synchrotron emissivity and synchrotron self-absorption coefficient are evaluated using the expression from Chiaberge & Ghisellini (1999); the synchrotron intensity as a solution of the radiative transfer equation for the case of spherical geometry is adopted from Katarzyński et al. (2001). The inverse Compton emission is computed following the approach by Katarzyński et al. (2001), which includes the full Klein–Nishina cross-section (Jones 1968). Finally, we transform the emission from the blob reference frame to the observer’s frame. We adopt a value for the Hubble constant of  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The absorption of  $\gamma$ -rays due to their interaction with the Extragalactic Background Light (EBL) is taken into account, with the use of a publicly available module.<sup>1</sup> For the current application, we use the EBL

<sup>1</sup><https://github.com/me-manu/ebtable>

model by Domínguez et al. (2011). The light curves are calculated by integrating over the time-dependent emission in the energy range of interest.

The treatment of the evolution of the electron spectrum was verified using test equations from Park & Petrosian (1996), as well as by comparing to analytical solutions for simple cases of the kinetic equation. The description of the radiative emission was cross-checked with the output of the code by Cerruti et al. (2015).

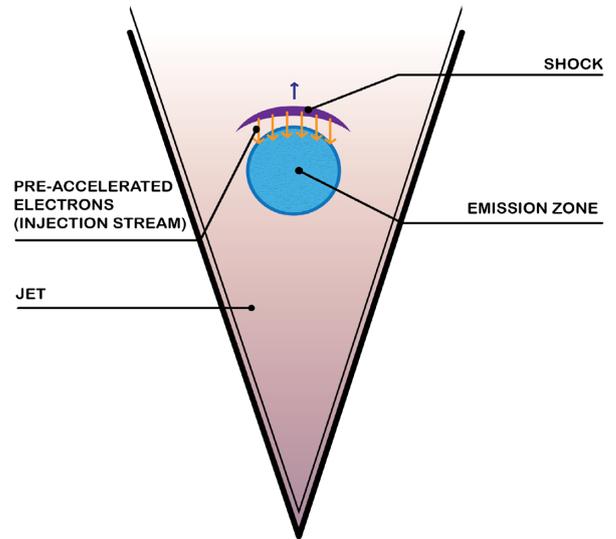
#### 4 A MODEL FOR THE LOW-STATE EMISSION OF MRK 421

In this paper, we consider that flaring activity is not simply superposed to the steady emission of the source, but caused by a perturbation of its quiescent (steady) low-state. This assumption results in significant additional constraints on the VHE emitting zone. Various effects can induce such a disturbance, including variable injection rate (e.g. Mastichiadis & Kirk 1997), a passing shock (e.g. Marscher & Gear 1985; Sikora et al. 2001), various instabilities (e.g. Meliani & Keppens 2009; Bromberg & Tchekhovskoy 2016), even stars crossing the jet (Barkov et al. 2012), etc. We will focus on scenarios where particle acceleration processes (Fermi-I and Fermi-II) are responsible for launching the flares. For this purpose, we first model the steady state of Mrk 421, and then incorporate perturbations and simulate the flare.

To describe the low state, we use the data set from Abdo et al. (2011), in which authors present the SED of Mrk 421 averaged over the observations taken during the MWL campaign from 2009 January 19 to June 1 when the source showed a very low level of activity. We consider that the low-state measurement by Abdo et al. (2011) represents the SED of the quiescent emission of the source on a relatively long-term time range. As the low-state observations and 2010 February flare are separated by only about 1 yr, we assume that the VHE blob did not undergo significant adiabatic expansion during this period and so that the broad-band emission of the source in the pre- and post-flare state in 2010 February is described by the low-state measurement by Abdo et al. (2011). The validity of this assumption is confirmed by the close match between the simulated quiescent source fluxes and observed pre- and post-flare fluxes in different energy bands.

We suppose that pre-accelerated electrons are continuously injected into the emission zone in the form of a steady ‘stream’ (see Fig. 1). Taking into account the analysis and conclusion by Yan et al. (2013) on the low-state of Mrk 421, we assume that the spectrum of the injected electrons  $Q_{\text{inj},0}(\gamma)$  is a power law with an exponential cutoff, resulting from a shock acceleration process. We also suppose that the electrons are injected only above a certain Lorentz factor  $\gamma_{\text{min,inj}}$ .

We simulate the steady state of the source with our EMBLEM code assuming an escape time-scale of  $t_{\text{esc}} = 1R_b/c$  from the blob. The population of electrons in the emitting region radiates and cools in accordance with the SSC scenario. The low state of Mrk 421 in our description is the asymptotically established equilibrium between the gain processes (injection) and losses (cooling and escape), and corresponds to the stationary solution of the kinetic equation (equation 1),  $\frac{\partial N_e}{\partial t} = 0$ . The blob radius  $R_b$  and Doppler factor  $\delta_b$  are constrained in a way that the related variability time-scale of the source derived from the causality arguments,  $t_{\text{var}} \sim R_b(1+z)(c\delta_b)^{-1}$ , is of the order of 1 d. This value is chosen since the 2010 February flare, considered to be a perturbation of the steady state, proceeds at a similar time-scale. Also, the variability time-scale of 1 d is used by Abdo et al.



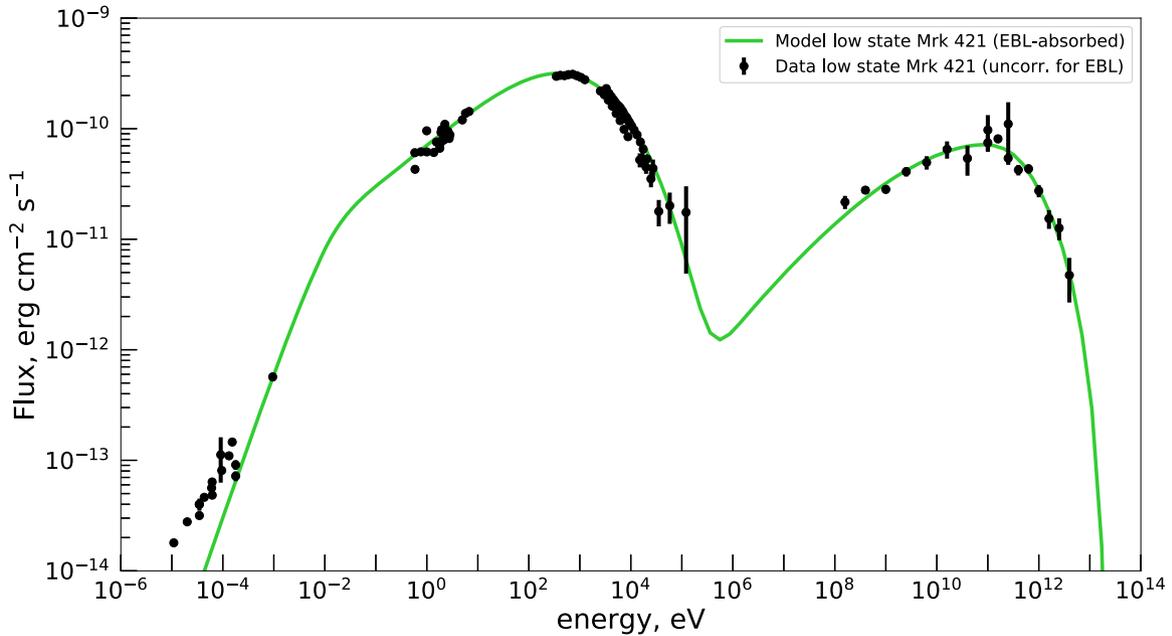
**Figure 1.** Scheme illustrating a physical scenario for the long-term low-state emission of Mrk 421. The blue filled circle represents the VHE  $\gamma$ -ray emitting zone (the blob) moving along the jet. The violet curve indicates the stationary shock leading the blob, which continuously accelerates particles of the upstream plasma and injects them into the downstream emission region. This injection flux is displayed with the orange arrows. The continuous influx of electrons pre-accelerated by the Fermi-I mechanism is responsible for the long-term quiescent emission of the source. Particles injected into the emitting zone radiate and cool in accordance with the SSC scenario, and escape from the blob at a time-scale  $t_{\text{esc}} = 1R_b/c$ .

(2011) in their instantaneous modelling of the low-state data set. We vary physical parameters of the source until we satisfactorily fit the MWL data set (cf. Table 1). The parameters are consistent with the results of the instantaneous modelling presented by Abdo et al. (2011). The characteristic variability time-scale in our model is  $t_{\text{var}} \approx 0.4$  d. Fig. 2 displays the MWL measurements from Abdo et al. (2011) together with our steady-state model, which shows a very good agreement with the data. The observed radio emission in the energy range  $10^{-5} \leq E_\gamma \leq 10^{-3}$  eV is thought to be dominated by the synchrotron emission of the extended jet, which is not included in our radiative model.

It should be noted that the assumed injected electrons can be accelerated by a relativistic electron–positron shock as the inferred slope of their spectrum,  $\alpha_{\text{inj}} = 2.23$ , is very close to the theoretical predictions for such type of shocks (e.g. Sironi, Keshet & Lemoine 2015). The most natural solution is therefore to consider that the injection flow is due to a stationary relativistic shock in front of the blob, generated by the interaction of the upstream extended jet plasma with the blob (see Fig. 1). In such a type of scenario (Kirk, Rieger & Mastichiadis 1998), the particles of the upstream plasma would be passing through the shock, undergoing Fermi-I acceleration, and being injected downstream into the emitting blob. The cutoff at high energies in the injection spectrum may result either from the limited power of the shock accelerator (Hillas criterion for the maximal attained energy of a particle), or from a drop in efficiency of shock acceleration of particles with Lorentz factor above  $\gamma_{\text{cut}}$  when their Larmor radius becomes larger than the characteristic size of turbulent eddies in the medium downstream of the shock. The injection of electrons only above  $\gamma_{\text{min,inj}} = 800$  can be explained by their pre-acceleration to these Lorentz factors prior to the shock, e.g. in the vicinity of the central engine.

**Table 1.** Physical parameters of the Mrk 421 low state. 3rd column: our scenario, 4th column: instantaneous modelling by Abdo et al. (2011).

Parameters of quiescent state	Symbol	Our model	Abdo et al. (2011)
Magnetic field (G)	$B$	0.04	0.038
Comoving blob radius (cm)	$R_b$	$2.8 \times 10^{16}$	$5.2 \times 10^{16}$
Doppler factor	$\delta_b$	29	21
Time-scale of electron escape	$t_{esc}$	$1R_b/c$	Not defined
Spectrum of injected electrons	$Q_{inj,0}(\gamma)$	$A_{inj}\gamma^{-\alpha_{inj}} \exp(-\gamma/\gamma_{cut})$ , for $\gamma \geq \gamma_{min,inj}$ 0, for $\gamma < \gamma_{min,inj}$	– Not defined
Injection spectrum normalization ( $\text{cm}^{-3} \text{s}^{-1}$ )	$A_{inj}$	$2.63 \times 10^{-3}$	–
Injection spectrum slope	$\alpha_{inj}$	2.23	–
Min. Lorentz Factor in inj. spectrum	$\gamma_{min,inj}$	800	–
Cutoff Lorentz Factor in inj. spectrum	$\gamma_{cut}$	$5.8 \times 10^5$	–

**Figure 2.** Modelling of the spectral measurement of the long-term low-state emission of Mrk 421. The green curve represents the model – SED in the asymptotically established steady state with EBL absorption included (Domínguez et al. 2011), and black points show the MWL data set from Abdo et al. (2011). The observed  $\nu F_\nu$  flux in the VHE range is uncorrected for EBL absorption. The host galaxy contribution has been subtracted from the optical measurements, the data in the optical-to-X-ray range were corrected for the effect of Galactic extinction.

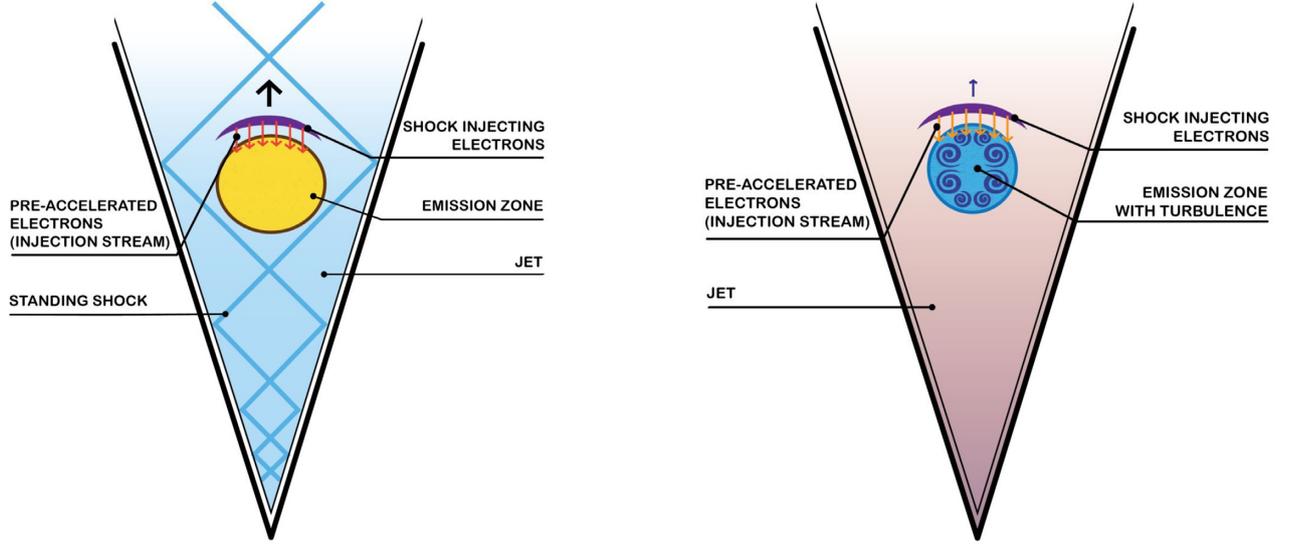
## 5 A GENERAL VALIDITY CRITERION FOR ONE-ZONE FLARING SCENARIOS WITH A TRANSIENT SHOCK

To describe the flare emission, a one-zone model is the most basic scenario that can be tested. A single emission region is then responsible for both the low-state emission and the short-term flaring emission, caused by a perturbation in the emitting blob. Strong flux rise in the VHE part of the spectrum can be the signature of enhanced particle acceleration while the emitting blob is disturbed by e.g. a passing shock and turbulence which induces a ‘boost’ or hardening of the steady-state particle spectrum. In the rest of this section, we focus in detail on the scenario in which the flare is induced by a shock passing through the emitting blob (Fig. 3, left-hand panel).

Following our main assumption, physical parameters of the emitting zone do not significantly change on average during the passage of the shock, namely magnetic field, size, Doppler factor, escape time, as well as injection function, and we neglect the inverse Compton

cooling so that the cooling rate is constant in time. The particles inside the blob are re-accelerated by the transient shock, the steady-state electron spectrum  $N_{e,0}(\gamma)$  is then perturbed and the blob emission as well. We show hereafter that when this scenario reproduces a flux increase observed at some frequency, for instance in the X-rays, the resulting flux induced at another frequency, for instance in the optical, can be or not perturbed just at the right level to fit the data, which can confirm or reject the model. We first need to connect the evolution of the electron spectrum to the flux variations for a generic case, then determine the general form of a time-dependent electron spectrum disturbed by a transient shock. Then using the derived general form of the electron spectrum we fit the X-ray data and find the shock acceleration time-scale, and as a final step we predict the optical flux increase to be compared to the observed one.

To explore how the photon flux enhancement is linked to the electron spectrum time evolution, we consider here only synchrotron emission variations (e.g. for Mrk 421 from radio band to hard X-



**Figure 3.** Sketches illustrating possible one-zone scenario for flares. The violet curve indicates the steady shock in front of the blob, which accelerates upstream particles and supplies the electron stream injected in the blob for the stationary low-state emission (shown in orange arrows). Left-hand panel: sketch in which the observed flux increase is due to a transient shock event perturbing the electron population in the emitting region. For instance, a standing shock with a so-called ‘diamond structure’ is present in the jet. The blob is passing through the knot of such a standing shock and the electron population is (re-)accelerated by transient Fermi-I process. Right-hand panel: sketch in which the quiescent blob is perturbed by turbulence induced for instance by extended jet inhomogeneities and the flare is triggered by stochastic acceleration of particles.

rays), using the  $\delta$ -approximation which considers that an electron with a Lorentz factor  $\gamma$  emits only at its critical frequency corresponding to the photon energy (in the observer’s frame)  $E_\gamma \simeq 5 \times 10^{-9} B_G \gamma^2 \delta_b (1+z)^{-1}$  (Rybicki & Lightman 1979), where  $B_G$  is the magnetic field in Gauss. The synchrotron SED is then given by (e.g. Dermer & Schlickeiser 2002)

$$E^2 \frac{dN_{\text{ph,syn}}}{dE}(E) \propto \bar{\gamma}^3 N_e(\bar{\gamma}), \quad (9)$$

where  $\bar{\gamma} = \sqrt{E/\kappa}$ , which is the Lorentz factor of an electron emitting synchrotron photon with energy  $E$ , and  $\kappa = 5 \times 10^{-9} B_G \delta_b (1+z)^{-1}$ .

The energy and the photon flux of a light curve in the energy range from  $E_{\text{min}}$  to  $E_{\text{max}}$  are given by

$$F_{\text{erg}} \propto \int_{E_{\text{min}}}^{E_{\text{max}}} E \frac{dN_{\text{ph,syn}}}{dE}(E) dE \propto \int_{E_{\text{min}}}^{E_{\text{max}}} E^{1/2} N_e(\bar{\gamma}(E)) dE$$

$$F_{\text{ph}} \propto \int_{E_{\text{min}}}^{E_{\text{max}}} \frac{dN_{\text{ph,syn}}}{dE}(E) dE \propto \int_{E_{\text{min}}}^{E_{\text{max}}} E^{-1/2} N_e(\bar{\gamma}(E)) dE.$$

These expressions connect the electron spectrum variations to those of the energy and photon flux in a certain energy range. Using them, we write out the flux increase factor  $\xi_{\text{LC}}$ , ratio of the peak flux in the light curve  $F_{\text{peak}}$  to the quiescent flux  $F_0$ ,

$$\xi_{\text{LC}} \equiv \frac{F_{\text{peak}}}{F_0} = \frac{\int_{E_{\text{min}}}^{E_{\text{max}}} E^s N_{e,\text{peak}}(\bar{\gamma}(E)) dE}{\int_{E_{\text{min}}}^{E_{\text{max}}} E^s N_{e,0}(\bar{\gamma}(E)) dE}, \quad (10)$$

where  $s = 1/2$  for the energy flux ratio, and  $s = -1/2$  for the photon flux ratio.

For the spectral flux increase factor  $\xi_{\text{spec}}$  at a particular photon energy  $E$ , a similar expression can be written using the equation (9):

$$\xi_{\text{spec}} \equiv \frac{dN_{\text{ph,syn,peak}}(E)/dE}{dN_{\text{ph,syn,0}}(E)/dE} = \frac{N_{e,\text{peak}}(\bar{\gamma}(E))}{N_{e,0}(\bar{\gamma}(E))}. \quad (11)$$

To compute the flux increase factor, we now focus on the analytical derivation of the electron spectrum evolution during the passage of a shock through the emitting blob.  $N_{e,0}(\gamma)$  is the steady state electron

spectrum and  $N_{e,\text{FI}}(\gamma, t)$  the evolving one during the passage of the shock. The evolution of  $N_{e,\text{FI}}(\gamma, t)$  is governed by the kinetic equation with the transient shock acceleration term,

$$\frac{\partial N_{e,\text{FI}}(\gamma, t)}{\partial t} = \frac{\partial}{\partial \gamma} (W(\gamma) N_{e,\text{FI}}(\gamma, t)) - \frac{N_{e,\text{FI}}(\gamma, t)}{t_{\text{esc}}} + Q_{\text{inj}}(\gamma), \quad (12)$$

with  $W(\gamma) = b_c \gamma^2 - \gamma/t_{\text{FI}}$ .

This equation can be solved analytically as presented in the Appendix A, so that

$$N_{e,\text{FI}}(\gamma, t) = N_{e,0}(\gamma) + \int_0^t \frac{\Gamma(\gamma, t, t') \cdot \exp[(1/t_{\text{esc}} - 1/t_{\text{FI}})(t' - t)]}{b_c t_{\text{FI}} \gamma^2} \cdot [Q_{\text{inj}}(\Gamma(\gamma, t, t')) + (b_c \Gamma(\gamma, t, t') - 1/t_{\text{esc}}) N_{e,0}(\Gamma(\gamma, t, t'))] dt', \quad (13)$$

where

$$N_{e,0}(\gamma) = \frac{1}{b_c \gamma^2} \int_\gamma^{\gamma_{\text{max}}} Q_{\text{inj}}(\gamma') \cdot \exp\left(\frac{1/\gamma' - 1/\gamma}{b_c t_{\text{esc}}}\right) d\gamma' \quad (14)$$

and

$$\Gamma(\gamma, t, t') = \frac{\gamma \cdot e^{(t'-t)/t_{\text{FI}}}}{1 + \gamma b_c t_{\text{FI}} (e^{(t'-t)/t_{\text{FI}}} - 1)}. \quad (15)$$

The parameter  $\gamma_{\text{max}}$  in the equation (14) represents the maximal Lorentz factor of the electron population in the blob. As already discussed in sub-section 3.5, throughout this paper, we set the Lorentz factor range in numerical computations wide enough (with large margins), allowing particles to freely migrate between the energy bins when they gain or lose energy. Therefore, we calculate the steady-state electron spectrum given by the equation (14) with  $\gamma_{\text{max}} \rightarrow \infty$ .

Knowing how the electron spectrum evolves during the passage of the shock, we can now evaluate the expected flux increase at different wavelengths. First, we compute the steady-state electron spectrum from equation (14), using the fit of the steady-state SED

of the source which one should obtain using the approach described in Section 4 with the physical parameters (including  $B$ ,  $\delta_b$ ,  $t_{\text{esc}}$ ) and the injection function  $Q_{\text{inj}}(\gamma)$ . The duration during which the transient shock acceleration is active is given by the crossing time of the shock, which can be approximated by the observed rise time of the flare, corrected for relativistic effects, i.e.  $t_{\text{cs}} \approx t_{\text{rise}} \delta_b$ . We then find the electron spectrum at the flare peak (at the moment when the shock exits the blob) depending on the Fermi-I time-scale by evaluating  $N_{e,\text{FI}}(\gamma, t)$  (equation 13) at the moment  $t = t_{\text{cs}}$ :  $N_{e,\text{peak}}(\gamma; t_{\text{FI}}) = N_{e,\text{FI}}(\gamma, t = t_{\text{cs}}; t_{\text{FI}})$ , the semicolon indicates separation between the arguments (Lorentz factor, time) and the parameter (Fermi-I time-scale). The corresponding acceleration time-scale is then deduced from the observed synchrotron flux increase in the X-ray band using equation (10), in which we use the peak electron spectrum  $N_{e,\text{peak}}(\gamma; t_{\text{FI}})$  obtained just above.

We solve numerically the equation for  $t_{\text{FI}}$ , and obtain its value  $t_{\text{FI}}^*$  required to produce the observed enhancement of the flux in the X-ray light curve. In case X-ray spectral measurements are available for the flare peak, one could derive the Fermi-I time-scale using those data, solving numerically the equation (11), to avoid the need of integration over the photon energies.

Finally, with the retrieved acceleration time-scale and peak electron spectrum, one can deduce from equation (10) or equation (11) the expected flux increase in the optical band and compare it to the one reconstructed from the optical data (after subtraction of the host galaxy contribution). In case of a significant divergence between the two values, one needs to conclude that the one-zone model with a shock traversing the emitting blob cannot satisfactorily describe the MWL data set, and reject the scenario.

Indeed, it is sufficient to retrieve only one parameter,  $t_{\text{FI}}$ , to apply the criterion when one knows the physical parameters of the low state and the duration of the flux rise, and when the optical and X-ray fluxes at the very peak are available. Conversely when there is no precise information on the time of the flux increase, there are two unknown parameters and an additional relation is needed. This can be achieved for instance if the flux increase ratio is available in both the soft and the hard X-ray bands, since one can then deduce  $t_{\text{cs}}$  and  $t_{\text{FI}}$  by solving numerically a system of two equations, namely applying equation (10) with  $N_{e,\text{peak}}(\gamma) = N_{e,\text{FI}}(\gamma; t = t_{\text{cs}}, t_{\text{FI}})$  to the soft and to the hard X-ray fluxes. Then one can predict the factor of the optical flux increase by applying the two inferred parameters  $t_{\text{cs}}$  and  $t_{\text{FI}}$ .

## 6 SCENARIOS FOR THE FLARE EMISSION OF MRK 421

In this section, we explore whether the 2010 February flare of Mrk 421 can be interpreted as being due to a moderate and non-destructive perturbation of the quiescent VHE emission region of the source. One-zone scenarios which provide the simplest way to connect low-state emission to the flaring one are considered first. There are then only two main free parameters to describe the flaring state, the time-scale of the shock or stochastic acceleration processes  $t_{\text{FII}}$ , and the duration of the acceleration phase  $t_{\text{dur,FII}}$  related to the rise time of the flare, which strongly constrains the picture. Two-zone scenarios are then developed to better account for the complexity of the observed MWL time evolution.

### 6.1 One-zone models

We first attempt to fit the flare data set by perturbing the low-state with a crossing shock or with turbulence, strictly within the central blob

radiating the steady VHE emission. According to the assumption of weak perturbation, macroscopic physical parameters describing the steady state are kept constant at first order.

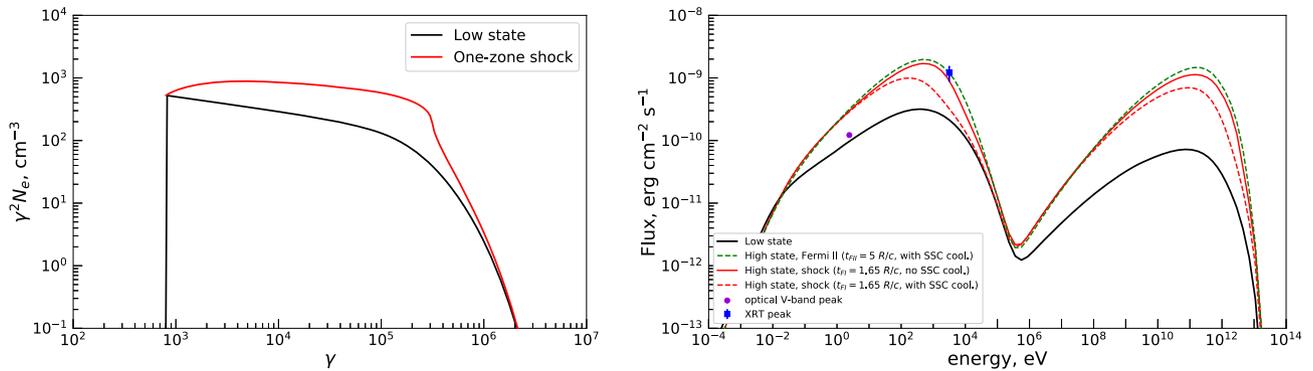
#### 6.1.1 Analytical results: a passing shock

In this case, the flare is simply initiated by a shock crossing the emitting blob, perturbing the particle population. The general criterion obtained in Section 5 can be directly applied considering the X-ray light curve in the energy range between 0.5 and 2 keV, and the optical V-band ‘light curve’ (host galaxy subtracted). The flux increase ratio for the X-ray light curve  $\xi_x = F_{x,\text{peak}}/F_{x,0} \approx 3.7$  and the rise time of the X-ray flare is around 3 to 4 d (in the observer’s frame). We therefore assume an average value  $t_{\text{cs}} \approx 101.5$  d, which translates into a shock speed relative to the blob of the order of  $\beta_{\text{sh}} \approx 0.1$ , and recover  $t_{\text{FI}} \approx 1.65 R_b/c \approx 17.8$  d from the equation (10) for the observed value of  $\xi_x$ . Indeed the final result on  $t_{\text{FI}}$  depends quite weakly on the duration of the shock passage.

Equation (10) now applied to the optical light curve shows that such passing shock induces a flux increase ratio  $\xi_{\text{opt}} \approx 3.3$  in the optical band, which is much higher than the observed value  $\xi_{\text{opt,obs}} \approx 1.26$ . Thus, the shock needed to reproduce the X-ray flare perturbs too much the optical flux, and one can conclude that this scenario is not satisfactory. This is illustrated in Fig. 4. The left-hand panel shows the electron spectrum, disturbed by the shock with Fermi-I acceleration and passage time-scales deduced from the X-rays. We verified that the analytical calculation of the electron spectrum (using equation 13) appears in close agreement with the numerical results obtained with the EMBLEM code. The right-hand panel of Fig. 4 displays the SED associated to the perturbed electron spectrum. One can clearly see that the shock causes too high pile-up in the optical part of the spectrum. Even though the analytical approach disregards the inverse Compton cooling, it does not affect the conclusion: a stronger shock needed to compensate the inverse Compton cooling and still describe the X-ray data, would lead to even higher rise of the optical flux. Therefore the criterion presented in Section 5 provides a lower limit on the optical flux increase.

#### 6.1.2 Numerical simulations results: shock and turbulence

While the analytical approach already excludes the one-zone shock acceleration model, one-zone models where Fermi-II acceleration plays a role need to be tested with the full numerical code. We explore various acceleration time-scales for models with only turbulent acceleration or with a combination of shock and turbulent acceleration. The duration of the acceleration phase  $t_{\text{dur,acc}}$  is dictated by the rise time of the light curve in the source frame, which is fixed from the 3.5 d as seen by the observer corrected by relativistic effects. In both cases, no combination of parameters succeed to produce a satisfactory fit of the data. For the case where only Fermi-II acceleration perturbs the blob, the one-zone model that provides a good representation of the X-ray flux increase while varying  $t_{\text{FII}}$  overshoots again the optical flux in the high state as can be seen in the right-hand panel of Fig. 4 (green curve). Contrarily, we find that the model describing well the optical peak data undershoots significantly the X-ray measurements. If we further consider that the escape time of electrons can be longer than  $1 R_b/c$  during the Fermi-II acceleration process (equations 6 and 7), it becomes even more difficult to achieve a good fit: the optical flux is even more overproduced. The problem with the optical flux excess persists when combining the shock and turbulent particle acceleration: we



**Figure 4.** Simulated electron spectrum and SEDs for one-zone flare scenario. Left-hand panel: analytical calculation (using equation 13) of the electron spectrum disturbed by a shock with  $t_{\text{FI}} = 1.65 R_b/c$  at the moment of the peak of the flare (red), compared to the steady-state electron spectrum (black). The inverse Compton cooling effect is not included for both spectra. Right-hand panel: peak SEDs for shock (dashed red curve) and Fermi-II (green curve) acceleration processes perturbing the blob simulated with the numerical code, with superimposed optical and X-ray spectral measurements near the flare peak. The black curve shows the steady-state SED. The solid red curve indicates the SED corresponding to the analytical electron spectrum shown in the left-hand panel (full SED computation, inverse Compton cooling neglected), dashed red curve represents the same as the solid one but with the inverse Compton cooling effect taken into account. The green curve displays the SED for the model in which Fermi-II process with  $t_{\text{FI}} = 5 R_b/c$  is acting on the blob population (inverse Compton cooling is taken into account). The Fermi-II time-scale is tuned in a way that the peak SED fits the X-ray data. For all the models, the acceleration process acts for 3.5 d in the observer’s frame ( $t_{\text{dur,FI/II}} = 101.5$  d in the frame of the blob). One can see that all one-zone models with shock or stochastic acceleration fitting the X-ray flare overproduce emission at the optical wavelengths.

did not find any set of the two acceleration time-scales  $t_{\text{FI}}$  and  $t_{\text{FII}}$  which mitigates the excess of the optical flux, and provides a reasonable description of the MWL peak data set. One-zone models with a moderate perturbation of the quiescent state appear too much constrained to explain the observed MWL flux variations during the flare.

## 6.2 Two-zone model

As one-zone models fail to satisfactorily describe the data set, we consider that the low-state and flaring emission emanate from two connected regions. The overproduction of the optical emission present in one-zone models can then be avoided if the optical emission is dominated by the quiescent region, and the X-ray flare by a second transient region. Since the conditions necessary for emission up to very high energies need to be fulfilled in both zones, their basic physical parameters cannot be too different. We assume that the steady-state emission comes from a relatively large region inside the extended jet as described in Section 4 and refer to it as the ‘quiescent blob’. The flaring emission originates from a smaller region that we identify as the ‘flaring region’, in contact with the quiescent blob. Both quiescent and flaring zones move relativistically along the jet with the same Doppler factor.

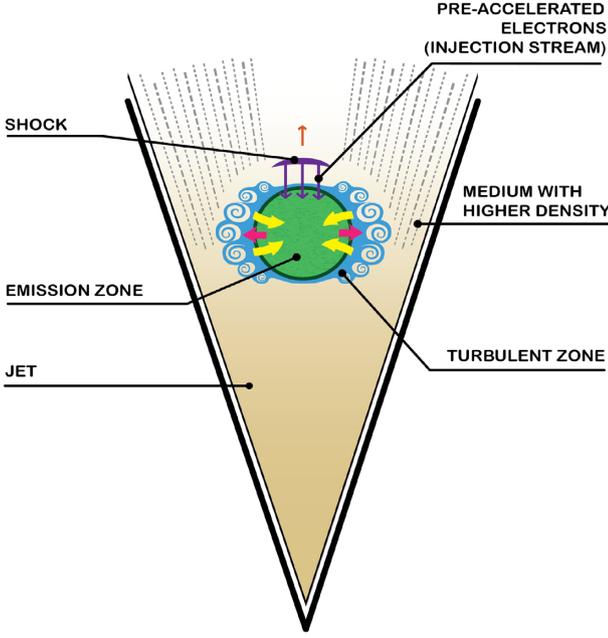
Yan et al. (2013) find that an instantaneous one-zone SSC scenario with a log-parabola electron spectrum provides a better fit of the Mrk 421 2010 February peak data than with a power-law electron spectrum. Their analysis suggests a turbulent re-acceleration process as the possible cause of the flare. Moreover, Zheng, Kang & Li (2014), invoking Fermi-II mechanism in their attempt to explain the 2010 February flare, conclude that the spectral and timing properties of the outburst are better described with the ‘hard-sphere’ approximation than with other turbulence types. This provides some support to the assumption of ‘hard-sphere’ turbulence ( $q=2$ ) adopted in sub-section 3.3 for our modelling of the Fermi-II acceleration effect. Considering these results, we focus our attention on the Fermi-II acceleration, sustained by ‘hard-sphere’ turbulence, as the process

powering activity in the flaring region. Because of that, we will also refer to the flaring region as ‘turbulent region’.

### 6.2.1 Application to the 2010 February flare

A generic configuration of the two-zone model we consider here is an abruptly appearing turbulent region at the interface of the quiescent blob and the surrounding jet. The two regions radiate and may exchange particles. There are two limiting cases: a steady-state emitting zone next to either (1) a non-radiative turbulent acceleration zone with an important particle escape, or (2) a radiative turbulent acceleration zone with negligible particle escape (see Fig. 5). In the first case, the typical size of the turbulent zone has to be comparable to the one of the quiescent emission, and the magnetic field should be much lower. In the second case the magnetic field should be commensurate to the one in the blob and the typical size has to be smaller. We consider the radiative contribution of the surrounding extended jet to be negligible above the radio band due, for instance, to significantly lower magnetic field and energetic particle density. After some trials, the second scenario proves to be the most promising one. Particles escaping from the quiescent blob are injected into the turbulent region, and are re-accelerated via the Fermi-II mechanism (see Fig. 5). The electrons in the turbulent region radiate a flaring synchrotron and IC emission and the observed flux increases.

Following our fundamental assumption interpreting the flare just as a small perturbation of the quiescent state, we apply to the quiescent blob our modelling of the Mrk 421 low state obtained in Section 4 with the physical parameters of Table 1, and consider that the transient turbulent region should admit rather similar values. However, there are now several new constraints and parameters for the physical description of the turbulent region. The effective size of the turbulent region,  $R_{\text{tr}}$ , is expected to be smaller than one of the blob,  $R_b$ , so that the additional inflow of particles from the turbulent region is less important than the steady injection inflow. In order to have a significant emission from the turbulent region, its typical magnetic field  $B_{\text{tr}}$  has to be sufficiently strong, i.e. comparable or higher to its strength inside the blob, and we first consider  $B_{\text{tr}} \sim$



**Figure 5.** Sketch showing a generic two-zone scenario for a flare in which a turbulent region is present around the quiescent emitting blob. The grey dashed lines show a material with higher density or different speed, which perturbs the medium around the blob and induces the turbulence. The violet curve above the quiescent blob represents the stationary shock wave, accelerating upstream electrons via Fermi-I process and injecting particles in the blob (injection flow indicated by violet arrows). The quiescent and turbulent regions exchange particles: ruby-coloured arrows display the injection of electrons escaping from the quiescent blob to the turbulent region, while yellow arrows represent the escaping particle flux from the turbulent region to the blob. The flow shown in yellow may be important or not (compared to the injection flow denoted by violet arrows), depending on the radii of the regions and on the respective escape time-scales. In the scenario we adopt in this paper, the particle flux from the turbulent region to the blob is negligible.

$B \sim 0.04$  G as in Table 1. The spectrum of particle injection in the turbulent region is defined by the spectrum of escaping particles from the quiescent blob, which is proportional to the steady-state spectrum of the electron population in the blob and assumed to be constant in time:

$$Q_{\text{inj,qr-tr}}(\gamma) \simeq \frac{N_{e,0}(\gamma)}{t_{\text{esc}}} \left( \frac{R_b}{R_{\text{tr}}} \right)^3 f_{\text{qr-tr}}, \quad (16)$$

where  $N_{e,0}(\gamma)$  is the steady-state spectrum (Section 4, and equation 14),  $t_{\text{esc}}$  is the escape time-scale from the quiescent emitting region (Table 1), and  $f_{\text{qr-tr}}$  is the fraction of particles escaping from the blob to the turbulent region.

The fraction  $f_{\text{qr-tr}}$  depends on the detailed geometrical configuration of the two zones and of the stationary front shock. The turbulent region forms a kind of thick turbulent torus around the quiescent blob. For practical use, we consider a simplified geometry and roughly describe the turbulent region as consisting of a few (namely 4 here) identical small spherical zones or ‘eddies’ at the lateral edge of the quiescent blob. One eddy has a radius  $R_{\text{ed}}$ , which under our assumptions is related to the effective size of the turbulent region  $R_{\text{tr}}$  via  $R_{\text{tr}}^3 = 4 R_{\text{ed}}^3$ . The contribution of each small eddy to  $f_{\text{qr-tr}}$  is determined by the solid angle  $\Omega$  enclosing it:

$$f_{\text{qr-tr}} = \frac{4 \cdot \Omega}{4\pi} \simeq \frac{1}{4\pi} 4 \frac{\pi R_{\text{ed}}^2}{R_b^2} = \left( \frac{R_{\text{ed}}}{R_b} \right)^2.$$

Plugging this relation in equation (16), one obtains:

$$Q_{\text{inj,qr-tr}}(\gamma) = \frac{N_{e,0}(\gamma) R_b}{4^{2/3} t_{\text{esc}} R_{\text{tr}}} = \frac{N_{e,0}(\gamma) R_b}{4 t_{\text{esc}} R_{\text{ed}}}. \quad (17)$$

The escape time-scale in the presence of turbulence is  $t_{\text{esc,ed}} \propto \delta B^2 / (2\mu_0)$  (see equation 8), and the Fermi-II acceleration time-scale is  $t_{\text{FII}} \propto \beta_A^{-2} (\delta B^2 / (2\mu_0))^{-1}$  (see equation 5). The Alfvén speed is time-dependent due to varying electron density in the turbulent region. A certain profile for the temporal evolution of the turbulence is needed to complete the description of the transient turbulent region,  $\delta B^2 = \delta B^2(t)$ .

The energy density contained in magnetic field fluctuations at a given time is determined by the balance between the injection of the turbulent energy in the flaring zone and its losses due to the work done on the acceleration of particles. Neglecting other losses such as damping of waves and the dependence of the losses on wavenumber, the equation governing the time evolution of the turbulent energy  $U_{\text{turb}}(t) = \delta B^2(t) / (2\mu_0)$  is (Burn 1975):

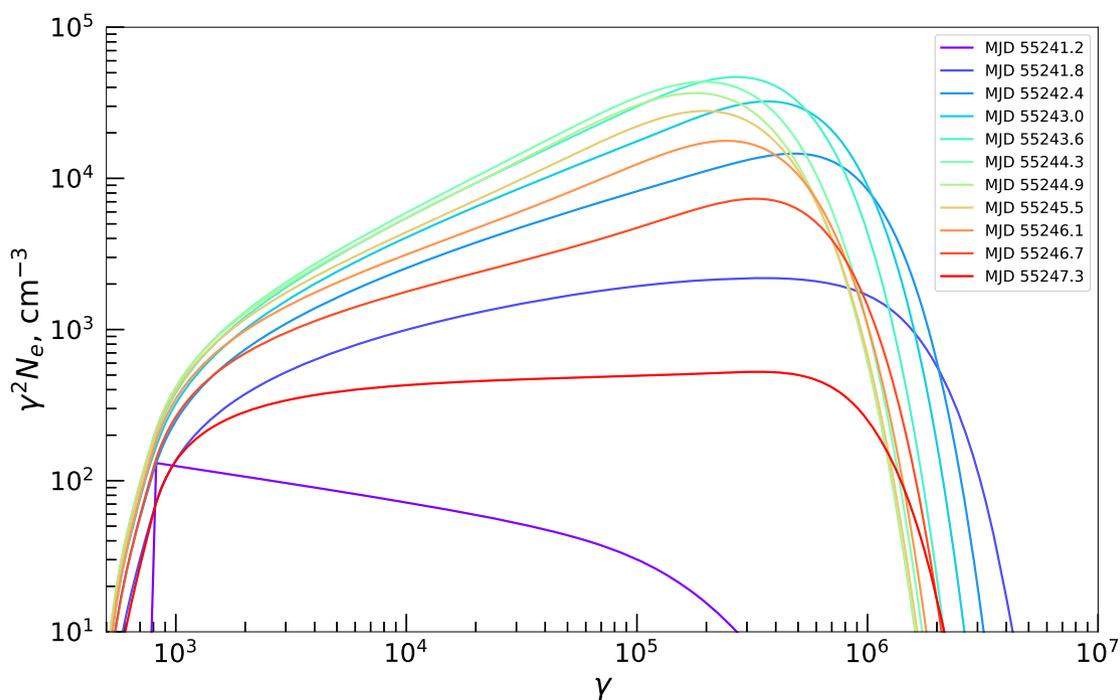
$$\begin{aligned} \frac{dU_{\text{turb}}(t)}{dt} &= Q_{\text{turb}}(t) - \int_{\gamma_{\text{min}}}^{\gamma_{\text{max}}} \frac{2\gamma m_e c^2}{t_{\text{FII}}(t)} N_{e,\text{tr}}(\gamma, t) d\gamma \\ &= Q_{\text{turb}}(t) - \frac{2\varepsilon(t)}{t_{\text{FII}}(t)}. \end{aligned}$$

Here  $Q_{\text{turb}}(t)$  is the time-dependent rate of turbulent energy injection,  $N_{e,\text{tr}}(\gamma, t)$  and  $\varepsilon$  are the electron spectrum and the energy density of particles in the turbulent region. The integral from  $\gamma_{\text{min}}$  to  $\gamma_{\text{max}}$  corresponds to the systematic energy gain by particles due to the Fermi-II acceleration process, evaluated by integration of the corresponding term in the kinetic equation (equation 1) and represents the loss term for the turbulence. From equations (5) and (6), this loss term writes as

$$\left( \frac{dU_{\text{turb}}(t)}{dt} \right)_{\text{loss}} \approx \frac{U_{\text{turb}}}{\lambda_{\text{max}}/c}.$$

The injection term  $Q_{\text{turb}}(t)$  depends on the detailed physics of the turbulence generation. For practical use we express this complex term with a minimal number of free parameters. The simplest scenario we first consider is close to a gate function with a continuous injection of a turbulence that is constant in time over a given duration. In this case, the energy density of the turbulence grows from zero to a constant maximum level on a time-scale of  $t_{\text{turb}} \approx \lambda_{\text{max}}/c$ . Once the injection is stopped, particles extract all available energy from the reservoir on the same time-scale, with the energy density falling off exponentially. This time-scale corresponds to the decay time of the longest mode in the turbulent spectrum, and is the shortest possible time for the build-up or dissipation of the turbulence. However such temporal profiles of the turbulent energy induce flares with an extended plateau and do not generate the observed 2010 February flare shape. To achieve a better representation of the data, we suppose a non-constant profile of the injection with linear rise and decline on time-scales  $t_{\text{turb,r}}$  and  $t_{\text{turb,d}}$  respectively, both much longer than  $\lambda_{\text{max}}/c$ , so that the temporal behaviour of the turbulent energy density  $\delta B^2(t) / (2\mu_0)$  approximately replicates the behaviour of the injection function  $Q_{\text{turb}}(t)$ . This situation can correspond, for instance, to the central quiescent blob traversing a dense region with a density gradient. We further assume that the turbulent energy density is about equal to the energy density of the non-turbulent large-scale component of the magnetic field at the peak of the flare,  $\delta B^2|_{\text{peak}} \sim B_{\text{tr}}^2$ .

We can now proceed to compute the emission of the turbulent region. We model the emission of one eddy, and then scale its flux by a factor of 4 to estimate the total emission. After the turbulence has dissipated, the region dissolves in the ambient medium inside



**Figure 6.** Simulated time evolution of the electron spectrum in the turbulent region during the flare. Electrons are injected into the turbulent region with Lorentz factors above  $\gamma_{\text{min,inj}} = 800$ , as they are initially injected into the quiescent blob above this Lorentz factor. The electron distribution then evolves from violet to red curves. The evolution is displayed with a time-step of  $\sim 0.6$  d.

the jet, which we simulate by simply stopping particle injection. For a self-consistent description we evaluate the Alfvén speed (from equation 6) and the escape and Fermi-II time-scales (from equation 8 and equation 5) at each time-step in the code. The escape of particles from the eddy at the beginning and at the end of the turbulence is considered as free-streaming with an escape time-scale  $t_{\text{esc,ed},0} = 1 R_{\text{ed}}/c$ , while during the turbulence phase the profile of the escape time-scale  $t_{\text{esc,ed}}(t)$  mimics the profile of the turbulent level. The acceleration time-scale profile is more complex due to its dependence on the inverse square of the varying Alfvén speed.

When dealing with two emission regions, one needs also to consider the possible contribution of external Compton (EC) processes: the flaring emission from the turbulent region is scattered off the relativistic electrons in the quiescent blob, and the steady-state emission from the blob off the relativistic electrons in the turbulent region. The former effect appears negligible for our conditions, but not the latter one. We thus use the sum of the synchrotron radiation of the flaring zone and of the quiescent blob as the seed photon field in the calculation of IC scattering on the electron population and of the IC cooling rate in the turbulent region. This leads to an average increase of flux in GeV-to-TeV  $\gamma$ -rays by  $\sim 40$  per cent and a decrease of the flux in soft-to-hard X-rays by a similar value, with respect to the case where this EC effect is not included. The effect is therefore rather significant.

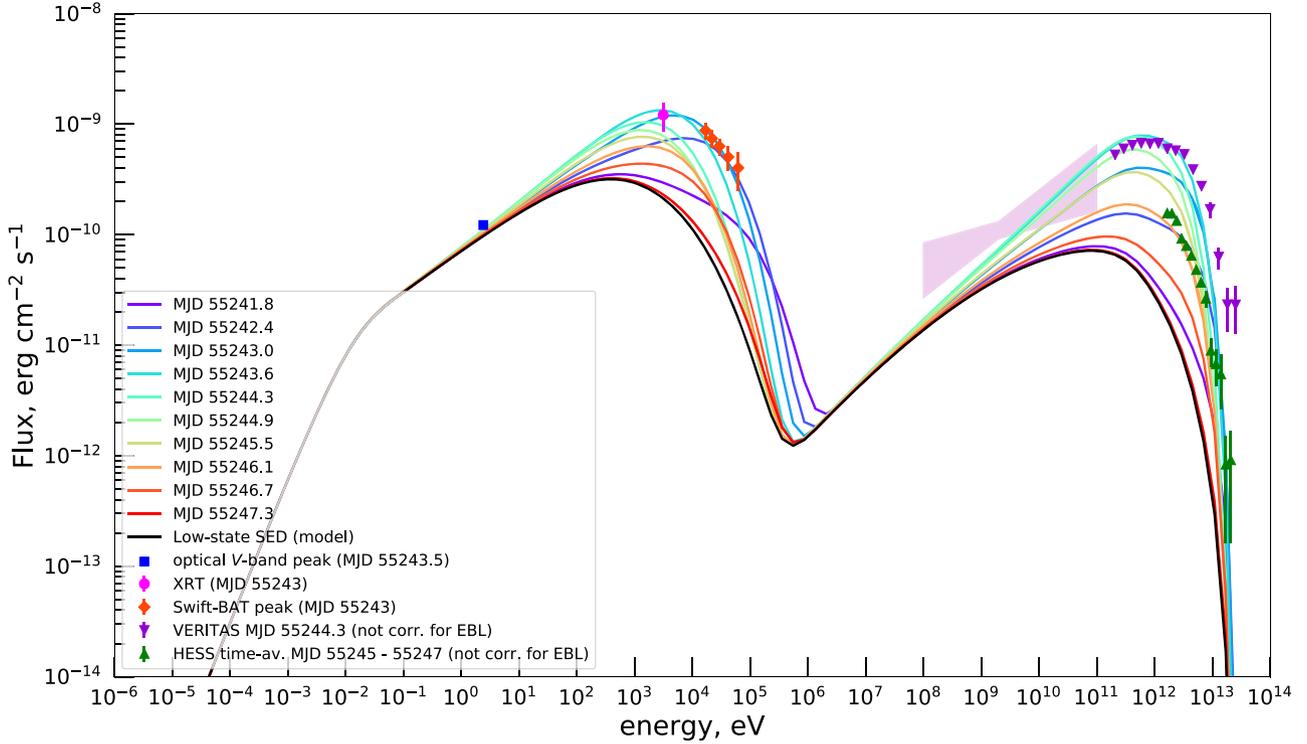
For the adjustment of the model to the data, we vary the five free parameters finally describing the flaring state, namely the magnetic field strength  $B_{\text{tr}}$  and the typical size of the turbulent region  $R_{\text{tr}}$ , the longest wavelength in the wave-turbulent spectrum  $\lambda_{\text{max}}$ , regulating the longest escape time-scale during the turbulence, and the rise and decay time-scales of the turbulent energy injection rate defined in the source frame,  $t_{\text{turb,r}}$  and  $t_{\text{turb,d}}$ , respectively. The sum of the emission of the quiescent blob and the time-dependent emission of the turbulent region are then compared to the observed MWL emission.

### 6.2.2 Results

The temporal evolution of the electron spectrum in the turbulent region is illustrated in Fig. 6. One clearly distinguishes the effect of particle acceleration, leading to an increase in the maximum electron energy and a hardening of the particle spectrum. Once acceleration becomes inefficient, particle cooling and escape result in the decrease of the maximum energy and in spectral softening.

The corresponding time-dependent SED of the total emission obtained for our best solution is shown in Fig. 7, with observational data superimposed. The overall agreement with the available spectra is very satisfactory. The two-zone model arrives at a good description of the flux increase in the X-ray and  $\gamma$ -ray bands, while the optical flux remains nearly unperturbed. Only the low-energy part of the *Fermi*-LAT spectrum during the flaring state appears slightly underestimated.

Fig. 8 shows the comparison between the simulated light curves and the MWL data in different energy bands. Ignoring the secondary peak that is visible in the X-ray band, the model provides a good representation of the observed flux variations, especially in the X-rays. The modelling accurately reproduces the position of the X-ray peaks with the occurrence of the soft X-ray flux maximum about one day later than in the hard X-rays. The appearance of a ‘shoulder’ following the peaks in the X-ray band is also a natural prediction of the turbulent acceleration scenario, as discussed below. In the optical band, while the flux variation has the right amplitude, there is a small offset between the observed and simulated flux amplitude at the level of  $\sim 10$  per cent. However a systematic shift may be present (i) in the optical data due to uncertainties in the host galaxy flux subtraction, and/or (ii) in the modelling of the low-state net optical flux (note the non-negligible scatter of the data in the optical band in Fig. 2). Therefore, given that we reproduce the observed variability amplitude, our description of the optical data can be considered as acceptable.



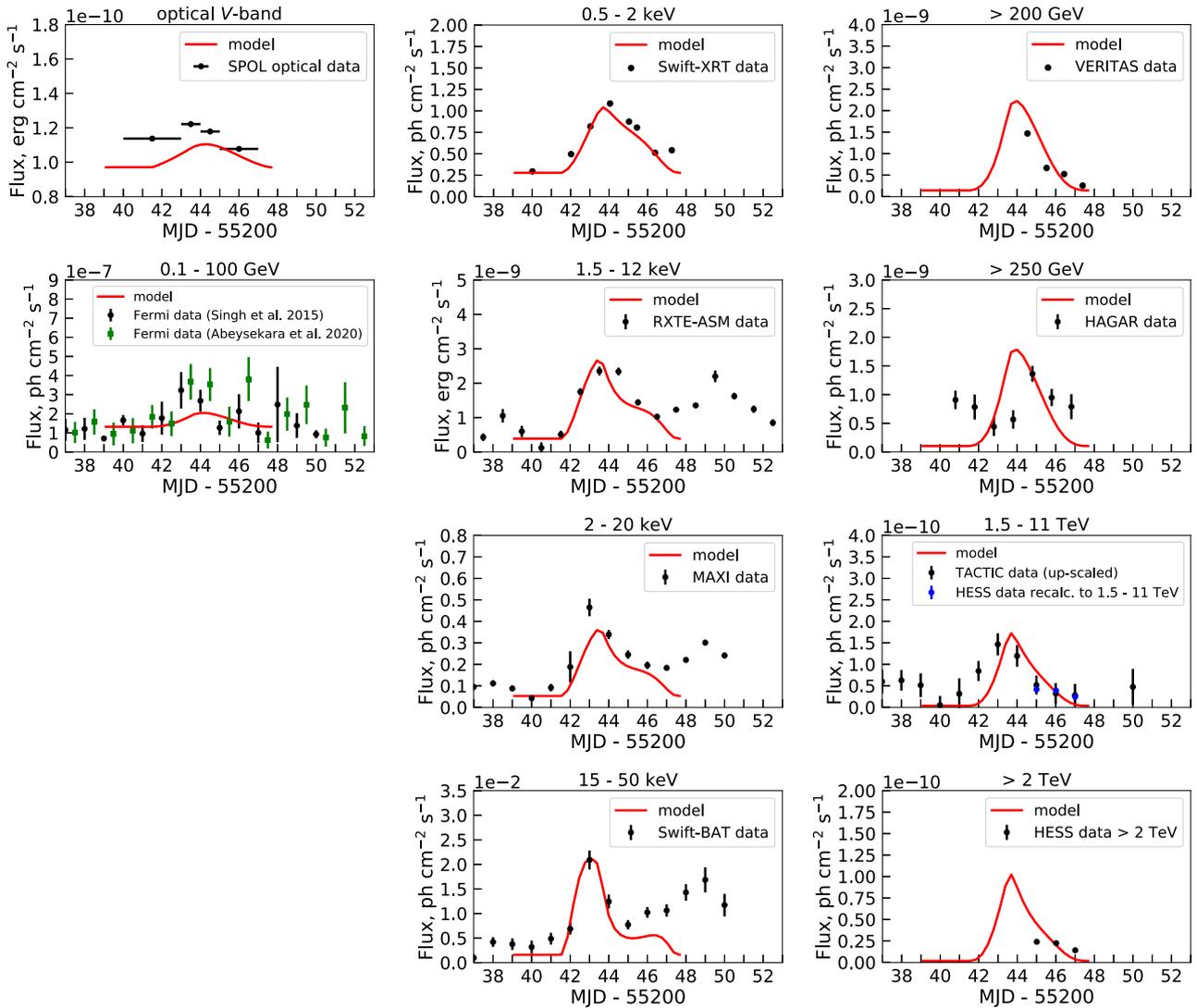
**Figure 7.** Simulated time evolution of the broad-band SED during the 2010 February flare compared to the spectral measurements from the flare data set. The modelled SEDs are absorbed on the EBL using the EBL model by Domínguez et al. (2011). The black line corresponds to the SED model of the low state of the source, and the SED then evolves from violet to red curves during the flare. The evolution is shown with a time-step of  $\sim 0.6$  d. The blue square point indicates the optical flux during the flare peak (host galaxy subtracted), the magenta round point the flux detected by XRT at  $\sim 3$  keV near the flare peak (2010 February 16, MJD 55243), the red diamond points the *Swift*-BAT peak SED (2010 February 16), the violet down-pointing triangle points the VERITAS SED measurement during 2010 February 17 (MJD 55244.3, not corrected for EBL), the green up-pointing triangle points the H.E.S.S. time-averaged SED over the period 2010 February 17–20 (MJD 55245.0–55247.0) during the flare decay (not corrected for EBL). The pink butterfly represents the *Fermi*-LAT SED uncertainty band for the spectral measurement during 17–20. Optical data (host galaxy subtracted) is taken from Shukla et al. (2012), VERITAS spectrum from Fortson et al. (2012), H.E.S.S. measurements from Tluczykont (2011). The data of XRT, *Swift*-BAT, and *Fermi*-LAT are adapted from Singh et al. (2015).

Numerical values of the parameters describing the turbulent region are provided in Table 2. The effective size of the turbulent zone appears to be an order of magnitude smaller than one of the quiescent blob, while the magnetic field in the turbulent region remains similar to the one in the quiescent blob. The maximal wavelength in the wave-turbulence spectrum cannot exceed the size of the turbulent region,  $\lambda_{\max} \leq R_{\text{tr}}$ , since modes of the turbulent cascade cannot stretch beyond the turbulent region and  $\lambda_{\max} > R_{\text{tr}}$  would imply superluminal escape of electrons (see equation 8). Our best-fitting model fulfils this condition with  $\lambda_{\max} \approx 0.023 R_{\text{tr}}$ , well below the size of the turbulent region. We can also verify that the flux of particles from the turbulent region to the quiescent one remains negligible. The relevant injection rate (per unit of volume) can be estimated as

$$Q_{\text{inj, tr-qr}}(\gamma, t) \simeq \frac{N_{\text{e, tr}}(\gamma, t)}{2 t_{\text{esc, ed}}(t)} \left( \frac{R_{\text{tr}}}{R_{\text{b}}} \right)^3 \quad (18)$$

assuming from geometrical arguments that half of the particles escaping the turbulent region reach the quiescent blob. We find that the spectrum of injected particles at the peak of the flare is an order of magnitude below the injection provided by the stationary front shock, therefore we conclude that the inflow of particles from the turbulent zone to the blob is indeed subdominant, and the emission from the quiescent blob remains approximately steady, in accordance with the proposed scenario.

The generic two-zone scenario reproduces in detail all the X-ray data related to the main flare (peaking at MJD 55243–55244), including its rise and decay temporal profiles, and the time lag of about 1 d observed between hard and soft X-rays. The time-dependent injection of turbulence and in particular the gradual decrease of the energy input induce a characteristic feature in the light curves, namely a flux stagnation before the end of the flare that is visible in the X-ray band and, to a lesser degree, at higher energies. This occurs due to a faster escape of particles when the level of the turbulence drops significantly. Because of the faster electron escape, the energy density in the turbulent zone decreases, the Alfvén speed increases, competing with the decreasing turbulence level, which leads to a temporary stabilization of the Fermi-II acceleration time-scale (see equation 5). The energy flux stalls, as the moderate acceleration balances cooling and escape. This continues until the fading of the turbulence becomes the dominant process and the stochastic acceleration time-scale quickly goes to infinity. Three X-ray light curves seem to show such a feature, before the secondary flare (peaking at MJD 55246–55247), although the presence of a second flare renders it difficult to establish. The model describes the behaviour of the X-ray energy flux, reaching the shoulder at that date. However the boost in Alfvén speed is not enough to initiate the observed secondary flare. Therefore, we argue that the secondary flaring event (which is seen only in X-rays, and not in  $\gamma$ -rays) might be caused by a ‘second wave’ of



**Figure 8.** Light curves generated by the two-zone model during the 2010 February flare compared to the MWL data. The observations include X-ray light curves by XRT and MAXI (Singh et al. 2015), *Swift*-XRT and RXTE-ASM (Shukla et al. 2012), the *Fermi*-LAT light curve computed by Singh et al. (2015) (with older instrument response functions) and by Abeysekara et al. (2020) (with newer instrument response functions), and the VHE light curves by H.E.S.S. (Tluczykont 2011), HAGAR (Shukla et al. 2012), TACTIC (recalibrated from Singh et al. 2015), and VERITAS (Abeysekara et al. 2020). The optical flux time evolution (host galaxy subtracted) is taken from Shukla et al. (2012).

**Table 2.** Physical parameters describing the flaring state and the turbulent region in the two-zone scenario.

Parameters for the 2010 February flare	Symbol	Value
Magnetic field (G)	$B_{\text{tr}}$	0.05
Comoving effective size of the turbulent zone (cm)	$R_{\text{tr}}$	$3.65 \times 10^{15}$
Maximal wavelength in the turbulent spectrum	$\lambda_{\text{max}}$	$0.023 R_{\text{tr}}$
Duration of the rise of the turbulent energy injection rate (source frame) (s)	$t_{\text{turb,r}}$	$5 \times 10^6$
Duration of the decay of the turbulent energy injection rate (source frame) (s)	$t_{\text{turb,d}}$	$10^7$

weaker turbulent energy injection, with another rise and decline of the turbulence injection rate, which we do not model here. The required secondary peak of injection could be caused by e.g. a second dense cloud on the way, or an echo-like growth of another instability mode.

As already mentioned, the adopted two-zone model slightly underproduces the *Fermi*-LAT spectra. One way to better describe the observed 0.1–1 GeV flux (Fig. 7) without perturbing its syn-

chrotron counterpart (optical flux) is to assume that there is an additional external Compton component. So far we have neglected the upscattering of flaring synchrotron emission off the quiescent blob electrons. With slightly different parameters of the turbulent region, this process could provide a more important contribution to the total GeV flux. In particular, a larger size of the turbulent zone would lead to a higher radiation density seen in the quiescent blob. In this case the particle flux from the flaring zone to the quiescent

emission region might become rather important and has to be taken into account. The *Fermi*–LAT light curve does not provide a tight constraint on the model.

At VHE, the two-zone model does not reproduce the exact shape of the VERITAS and H.E.S.S. light curves, which are the most constraining, but it appears compatible with the data set. The model fits reasonably well the VERITAS spectrum obtained at MJD 55244.3 (roughly 1 d after the flare peak), as well as the spectral measurement by H.E.S.S. averaged over time period MJD 55245.0–55247.0 (fall of the flare). Some discrepancy can be seen at the energies above  $\sim 5$  TeV. While our modelled SED may be slightly too soft at the highest energies, due to a relatively simple treatment of the turbulent acceleration and possibly systematics of the EBL model, it should be noted that the VERITAS spectrum points are still preliminary and may suffer from systematic uncertainties. Focusing on the night-to-night variations, we did not try to treat the intranight variability. One may consider that the fluctuations of the flux at the 1 h time-scale can be due to small flaring subregions in the emitting turbulent region. Indeed, we did not model the spatial structure of the turbulence region and approximated it for practical reasons with four independent spheres, while in reality it might be composed of many individual small cells having random direction of magnetic field and fluctuating velocity fields and electron density (Marscher 2014). Such more complex physics of the turbulent region could explain stochastic flux variations at short intranight time-scales.

## 7 DISCUSSION

The generic two-zone scenario developed in this paper provides one of the first fully time-dependent models able to reproduce reasonably well the evolution of the MWL flare despite a limited number of free parameters (Table 1 and 2), relying on the interpretation of the flare as a weak perturbation of the quiescent phase. The time evolution of the flare is dominated by two key parameters, the rising time of the turbulent energy injection rate  $t_{\text{turb,r}}$ , and the effective size of the turbulent zone  $R_{\text{tr}}$ . However, the uniqueness of the selected (imperfect) solution cannot be ensured.

### 7.1 The assumption of flare as a weak perturbation

Our initial hypothesis describing VHE flares as weak perturbations of the quiescent state appeared quite powerful to significantly constrain the flare models and to reduce the number of free parameters, but it should be tested by applying it to other flares once more detailed and complete MWL monitoring data sets become available. A preliminary analysis shows that the 2008 flare of Mrk 421 could be another example (Donnarumma et al. 2009). Indeed, AGN flares are just non-destructive transient events and, although remarkable during the outbursts, their total energy budget remains negligible compared to the energy radiated over years by the quiescent states. However, a limitation to this view could arise from a detection of extended VHE emission in blazars, unless most of the quiescent emission still comes from a compact radiatively dominant region. So far, extended VHE emission in AGN was detected only from the jet of a non-blazar source, the radio galaxy Cen A (H. E. S. S. Collaboration 2020). This result questions the validity of the present scenario for radio galaxies, which should be further investigated if VHE flares were detected from Cen A, which is not yet the case.

### 7.2 Alternative models

The one-zone model has served the SED modelling well in the past. However the situation becomes more challenging with increasing quality and completeness of data sets. One-zone scenarios for the 2010 February flare of Mrk 421 did not prove to be successful under our assumptions and several alternative two-zone scenarios can be considered. We developed here the two-zone scheme which appeared to us the most promising, based on previous works available in the literature on the 2010 February flare, but other options could be further analysed as well. As mentioned before, an alternative two-zone scenario can be provided by a non-radiative acceleration region around the emitting blob (Kirk et al. 1998), which injects particles, but contributes only a negligible amount of radiation due to a low magnetic field. The acceleration region would be injecting particles with a hard spectrum in the quiescent blob to launch the flare. Such a scenario should in principle be able to provide the inflow of electrons with a hard spectrum that is needed to arrive at a strong flux increase at high energies, while the optical flux remains relatively constant. However, as shown in Dmytriiev, Sol & Zech (2019) this scenario describes indeed well the varying synchrotron emission, but it underpredicts the  $\gamma$ -ray flux by a factor of  $\sim 3$  for the data set under study. A two-zone model considered by Cao & Wang (2013) in their modelling of the 2008 June flare of Mrk 421 assumes that the steady emission arises from the outer jet where particles are accelerated by the Fermi-I process and dominates the total flux in the radio-to-optical domain, while the variable component is produced in a much smaller inner jet region in which low-energy particles are injected continuously and undergo stochastic acceleration. The flare is then initiated by a change in the stochastic acceleration time-scale. Such a configuration is interesting, but the model still lacks a full time-dependent framework to self-consistently describe the SED evolution from the low state to the high one.

### 7.3 Complexity of the physics involved in the scenario

The turbulent acceleration of particles as the process driving flares of HBLs was previously considered by various authors. Different types of turbulence can be assumed. For instance, Tramacere et al. (2011) reproduce various trends observed in six HBL flares with stochastic acceleration of particles on a relatively short time-scales and conclude that two acceleration scenarios can provide a good description of the X-ray data: (i) the Fermi-II acceleration time-scale evolves because of variations of  $\delta B^2$  or  $\beta_A$ , while the ‘hard-sphere’ turbulence spectrum remains constant, or (ii) the index of the turbulence spectrum evolves with time. In this paper, we stick to the first scenario, considering however a more complex self-consistent evolution of the energy density of magnetic field fluctuations and of the Alfvén speed, as well as following a two-zone approach.

Several effects can lead to the generation of a turbulence in the vicinity of the emitting quiescent blob, assumed to stream at relativistic speed along the jet axis. One possibility is a sudden enhancement of plasma instabilities, or Kelvin–Helmholtz or rotationally induced Rayleigh–Taylor instability (Meliani & Keppens 2009) in the shear at the interface between the quiescent blob associated to a faster inner spine and the slower outer layer of the jet (Sol, Pelletier & Asseo 1989). The turbulence might also be excited as the blob is passing through a dense cloud of gas, resulting from e.g. an interaction between a red giant star and the jet (Barkov et al. 2012). Due to enhanced density of the plasma flowing past the blob, the Reynolds number could appear to be higher than the critical one, and as a result, a transient turbulent region is formed. The exact

physical conditions and plasma instabilities that trigger spontaneous and transient turbulence around the emitting blob question the broad field of turbulence generation and shock–turbulence interaction, and cannot be simulated in simple radiative models as developed here. As described above, we approximate the evolving injection rate of the turbulent energy with a simple linear rise and decay. However, we find that the exact shape of the injection function has only a small impact on the resulting shape of the light curves, as long as it represents a peak, characterized by suitable rise and fall time-scales. So, finally, a very rough modelling of the turbulence can still provide a reasonably good description of the observed event.

Clearly, all current available flare scenarios need to proceed with highly simplifying assumptions in view of the limited current knowledge on the VHE emitting regions, and of the expected underlying complexity of the physics involved. For instance, in the two-zone scenario developed here, one should probably expect some deformation of the ‘stationary’ front shock as a back reaction to the external perturbation supposed to induce the turbulence and the flaring at the edges of the quiescent blob, as well as a possible mutual shock–turbulence interaction (Andreopoulos, Agui & Briassulis 2000). A large number of linear and non-linear phenomena may occur which can potentially deform the stationary shock, impact on the quiescent emission, and modify the characteristics of the turbulence, which we completely neglected here. Instead of injecting directly the turbulence at the flanks of the central blob as a boundary effect, another possibility that we did not take into account could be as well to consider that the external perturbation acts first on the ‘quiescent’ front shock, which in turn can enhance the turbulence behind it and around the central blob. Magnetic reconnection could also play a role in such collision-less plasmas (Karimabadi et al. 2014; Nishikawa et al. 2020). However this process is not expected to contribute significantly to particle acceleration in our scenario involving a weakly magnetized shock.

#### 7.4 A new tentative interpretation of lognormality and different types of noises in blazars

In the two-zone scenario developed here, turbulence effects underlie both the quiescent emission and the flaring one, in a slightly different way through Fermi-I and Fermi-II mechanisms. In such a framework, it seems possible to attribute the lognormality observed in VHE light curves of some bright AGN, such as PKS 2155-304 (Aharonian et al. 2007; H. E. S. S. Collaboration 2010, 2017), to a turbulent process somewhat similar to the universal one suspected in laboratory turbulent flows (Mouri, Hori & Takaoka 2009), and to which the multiplicative central-limit theorem can be applied. Indeed, particle-in-cell simulations have also put in evidence a lognormal distribution of the particle and internal energy density in turbulent collisionless, magnetized, relativistic electron–positron plasmas similar to those expected in some AGN jets (Zhdankin et al. 2018). Under such interpretation, detailed studies of the power spectral density (PSD) could help characterizing the turbulence properties directly in the VHE emitting regions.

The measured PSDs appear as a power law  $P(\nu_i) \propto \nu_i^{-\beta}$  where  $\nu_i$  is the temporal frequency, with a variable index  $\beta$  in the range  $1 \leq \beta \leq 2$ . This reveals how the amplitudes of the variability are spread over the different time-scales and suggests that stochastic underlying processes are at work in the emitting regions, with correlated coloured noises typically of the flicker or pink type ( $\beta = 1$ ), or of the random walk or red type ( $\beta = 2$ ). Different values of the  $\beta$  index have been found for the quiescent and the flaring states in the two blazars PKS 2155-304 and in Mrk 421, with the same trends. They suggest

a pink noise for the quiescent VHE states and a red noise for the VHE flaring states, with  $\beta = 1.1_{-0.13}^{+0.10}$  on time-scales larger than one day to several years for PKS 2155-304 (H. E. S. S. Collaboration 2017) and  $\beta = 1.1_{-0.5}^{+0.5}$  for Mrk 421 on time-scales from months to years (Goyal 2020), and with  $\beta \simeq 2$  for PKS 2155-304 on time-scales from minutes to a few hours (Aharonian et al. 2007) and  $\beta \simeq 1.75$  for Mrk 421 on time-scales from seconds to hours during the night of the flare on 2010 February 17 (Abeysekara et al. 2020). In both sources the PSD is flatter for quiescent states and steeper for flares. A possible interpretation of these results, in the frame of the present two-zone model, suggests that the different acceleration mechanisms and underlying turbulences during the quiescent state or during the flares are at the origin of the different types of noises currently observed. Slowly variable shock acceleration (and particle injection) by the shock at the front of the blob, and long-term slowly varying turbulence inside the blob (with  $q \leq 2$ ), allow a better spread of the amplitudes of the variability over the different time-scales above 1 d, considering that long-term perturbations of the front shock can occur on time-scales of days, months, and years. Conversely, transient Fermi-II acceleration with  $q = 2$ , for the ‘hard-sphere’ turbulence, likely stores larger amplitudes of the variability at the time-scales of hours (corresponding to the largest spatial scales of the entire turbulent zone) compared to the smallest time-scales (minutes or seconds) due to turbulent cascade phenomena, resulting in a redder noise than for the quiescent long-term emission.

#### 7.5 Analogy with hotspots of extragalactic radio sources

As a final remark, we would like to emphasize that the complex configuration adopted for the generic scenario with two zones and both Fermi-I and Fermi-II mechanisms, was just reached step by step, starting from the simplest scenario and adding complexity only when needed in view of the constraints imposed by the data. Unexpectedly, such a circumstance has some similarities with the situation observed in hotspots of nearby radio galaxies like 3C 445, 3C 105, 3C 227, and 3C 195, for which different authors advocate multizone models for particle acceleration, involving both Fermi-I and Fermi-II mechanisms to describe the hotspot physics with a compact front shock and a more diffuse and turbulent region in its wake (Kruells 1992; Prieto, Brunetti & Mack 2002; Fan et al. 2008; Orienti et al. 2012; Isobe et al. 2017; Orienti et al. 2017; Migliori et al. 2020; Orienti et al. 2020). So, finally, the scenario built independently in this paper for the complex quiescent and flaring VHE emitting zone of Mrk 421 describes it as a kind of ‘mini hotspot’, appearing much earlier along the jet. Although compactness, energies, and temporal evolution involved are quite different, this analogy could provide some clues to better describe the VHE zone from information gathered at the larger and resolved scale of the hotspots.

## 8 CONCLUSIONS

We have provided a general analytical approach to determine the feasibility of a one-zone shock model to produce flaring events in blazars.

It was shown that the MWL data set of the 2010 February flare of Mrk 421 cannot be described with a one-zone shock model, and neither with turbulent acceleration in a one-zone model.

A self-consistent two-zone model, with a large emission region responsible for the steady-state emission and a smaller, connected turbulent region responsible for particle acceleration and emission during the flare, provides a very satisfactory description of the available MWL spectra and light curves for this event. The observed

spectral hardening and asymmetric flare profile are a direct outcome of the simulated acceleration, cooling, and particle escape processes.

In this model, the steady-state emission and flare emission can be connected with a limited number of free parameters, and the flare arises naturally as a perturbation due to a transient turbulence on the edge of the steady-state emission region.

In general, the scenario we present is a viable model for producing flares on a day time-scale. A future application to further flare data sets should show whether the proposed intermittent turbulent acceleration provides an interesting scheme to explain most blazar flares.

## ACKNOWLEDGEMENTS

The authors wish to thank M. Lemoine for useful discussions, C. Nigro for his permission to share his numerical implementation of the Chang and Cooper scheme (without acceleration), M. Meyer for the public availability of his PYTHON code to model a variety of EBL models. We also acknowledge S. Buriak for assistance in designing the sketches representing the flare scenarios. Finally, we thank J. Finke for providing us the data points of the Mrk 421 low state, K. K. Singh for providing the data points of the spectral measurements in different energy bands and MWL light curves during the Mrk 421 2010 February flare (presented in his 2014 paper) and M. Tluczykont for the H.E.S.S. data points measured during the flare.

## DATA AVAILABILITY

The data sets used in this paper were derived from articles in the public domain (references presented in Section 2), and will be shared on reasonable request to the corresponding author. The numerical EMBLEM code used for the data modelling cannot be shared publicly due to its ongoing development and extension. More information on the code can be obtained on reasonable request to the corresponding author.

## REFERENCES

Abdo A. A. et al., 2010, *Nature*, 463, 919  
 Abdo A. A. et al., 2011, *ApJ*, 736, 131  
 Abeysekera A. U. et al., 2020, *ApJ*, 890, 97  
 Acciari V. A. et al., 2014, *Astropart. Phys.*, 54, 1  
 Aharonian F. et al., 2007, *ApJ*, 664, L71  
 Ahnen M. L. et al., 2016, *A&A*, 593, A91  
 Aleksić J. et al., 2015, *A&A*, 578, A22  
 Andreopoulos Y., Agui J. H., Briassulis G., 2000, *Annu. Rev. Fluid Mech.*, 32, 309  
 Asano K., Hayashida M., 2018, *ApJ*, 861, 31  
 Banerjee B., Joshi M., Majumdar P., Williamson K. E., Jorstad S. G., Marscher A. P., 2019, *MNRAS*, 487, 845  
 Barkov M. V., Aharonian F. A., Bogovalov S. V., Kelner S. R., Khangulyan D., 2012, *ApJ*, 749, 119  
 Becker P. A., Le T., Dermer C. D., 2006, *ApJ*, 647, 539  
 Bell A. R., 1978, *MNRAS*, 182, 147  
 Böttcher M., Baring M. G., 2019, *ApJ*, 887, 133  
 Boutelier T., Henri G., Petrucci P. O., 2008, *MNRAS*, 390, L73  
 Bromberg O., Tchekhovskoy A., 2016, *MNRAS*, 456, 1739  
 Burn B. J., 1975, *A&A*, 45, 435  
 Cao G., Wang J., 2013, *PASJ*, 65, 109  
 Casadio C. et al., 2015, *ApJ*, 813, 51  
 Cerruti M., Zech A., Boisson C., Inoue S., 2015, *MNRAS*, 448, 910  
 Chang J., Cooper G., 1970, *J. Comput. Phys.*, 6, 1  
 Chiaberge M., Ghisellini G., 1999, *MNRAS*, 306, 551

Chitnis V. R., Pendharkar J. K., Bose D., Agrawal V. K., Rao A. R., Misra R., 2009, *ApJ*, 698, 1207  
 de Gouveia Dal Pino E. M., Piovezan P., Kadowaki L., Kowal G., Lazarian A., 2010, *Highlights Astron.*, 15, 247  
 Dermer C. D., Schlickeiser R., 2002, *ApJ*, 575, 667  
 Dermer C. D., Miller J. A., Li H., 1996, *ApJ*, 456, 106  
 Dmytriiev A., Sol H., Zech A., 2019, in 36th International Cosmic Ray Conference (ICRC2019), Volume 358. Madison, WI, U.S.A., p. 662  
 Domínguez A. et al., 2011, *MNRAS*, 410, 2556  
 Donnarumma I. et al., 2009, *ApJ*, 691, L13  
 Fan Z.-H., Liu S., Wang J.-M., Fryer C. L., Li H., 2008, *ApJ*, 673, L139  
 Fortson L., VERITAS Collaboration, Fermi-LAT Collaborators, 2012, in Aharonian F. A., Hofmann W., Rieger F. M., eds, AIP Conf. Proc. Vol. 1505, High Energy Gamma-Ray Astronomy: 5th International Meeting on High Energy Gamma-Ray Astronomy. Am. Inst. Phys., New York, p. 514  
 Gedalin M., 1993, *Phys. Rev. E*, 47, 4354  
 Giannios D., Uzdensky D. A., Begelman M. C., 2009, *MNRAS*, 395, L29  
 Goyal A., 2020, *MNRAS*, 494, 3432  
 Grimm H. J., Gilfanov M., Sunyaev R., 2002, *A&A*, 391, 923  
 H. E. S. S. Collaboration, 2010, *A&A*, 520, A83  
 H. E. S. S. Collaboration, 2017, *A&A*, 598, A39  
 H. E. S. S. Collaboration, 2020, *Nature*, 582, 356  
 Hovatta T. et al., 2015, *MNRAS*, 448, 3121  
 Isobe N., Koyama S., Kino M., Wada T., Nakagawa T., Matsuhara H., Niinuma K., Tashiro M., 2017, *ApJ*, 850, 193  
 Jaekel U., Schlickeiser R., 1992, *J. Phys. G: Nucl. Part. Phys.*, 18, 1089  
 Jones F. C., 1968, *Phys. Rev.*, 167, 1159  
 Joshi M., Böttcher M., 2011, *ApJ*, 727, 21  
 Kardashev N. S., 1962, *Sov. Astron.*, 6, 317  
 Karimabadi H. et al., 2014, *Phys. Plasmas*, 21, 062308  
 Katarzyński K., Sol H., Kus A., 2001, *Astron. Astrophys.*, 367, 809  
 Kirk J. G., Rieger F. M., Mastichiadis A., 1998, *A&A*, 333, 452  
 Kruells W. M., 1992, *A&A*, 260, 49  
 Larionov V. M. et al., 2016, *MNRAS*, 461, 3047  
 MAGIC collaboration, 2020, *ApJS*, 248, 29  
 Marscher A. P., 2014, *ApJ*, 780, 87  
 Marscher A. P., Gear W. K., 1985, *ApJ*, 298, 114  
 Mastichiadis A., Kirk J. G., 1997, *A&A*, 320, 19  
 Meliani Z., Keppens R., 2009, *ApJ*, 705, 1594  
 Migliori G., Orienti M., Coccato L., Brunetti G., D'Ammando F., Mack K. H., Prieto M. A., 2020, *MNRAS*, 495, 1593  
 Moderski R., Sikora M., Coppi P. S., Aharonian F., 2005, *MNRAS*, 363, 954  
 Mouri H., Hori A., Takaoka M., 2009, *Phys. Fluids*, 21, 065107  
 Nishikawa K. et al., 2020, *MNRAS*, 493, 2652  
 O'Sullivan S., Reville B., Taylor A. M., 2009, *MNRAS*, 400, 248  
 Orienti M., Prieto M. A., Brunetti G., Mack K. H., Massaro F., Harris D. E., 2012, *MNRAS*, 419, 2338  
 Orienti M., Brunetti G., Nagai H., Paladino R., Mack K. H., Prieto M. A., 2017, *MNRAS*, 469, L123  
 Orienti M., Migliori G., Brunetti G., Nagai H., D'Ammando F., Mack K. H., Prieto M. A., 2020, *MNRAS*, 494, 2244  
 Park B. T., Petrosian V., 1996, *ApJs*, 103, 255  
 Prieto M. A., Brunetti G., Mack K.-H., 2002, *Science*, 298, 193  
 Raiteri C. M. et al., 2017, *Nature*, 552, 374  
 Rybicki G. B., Lightman A. P., 1979, *Radiative Processes in Astrophysics*. Wiley-VCH, Weinheim  
 Schlickeiser R., 1989, *ApJ*, 336, 243  
 Shukla A. et al., 2012, *Astron. Astrophys.*, 541, A140  
 Shukla A. et al., 2018, *ApJ*, 854, L26  
 Shukla A., Mannheim K., 2020, *Nat. Commun.*, 11, 4176  
 Shukla A., Mannheim K., Chitnis V. R., Roy J., Acharya B. S., Dorner D., Hughes G., Biland A., 2016, *ApJ*, 832, 177  
 Sikora M., Błażejowski M., Begelman M. C., Moderski R., 2001, *ApJ*, 554, 1  
 Singh K. K. et al., 2017, *Nat. Astron.*, 54, 24  
 Singh K. K., Yadav K. K., Chandra P., Sahayanathan S., Bhatt N., Rannot R. C., Tickoo A. K., Koul R., 2015, *Astropart. Phys.*, 61, 32

- Sironi L., Keshet U., Lemoine M., 2015, *Space Sci. Rev.*, 191, 519  
 Skilling J., 1975, *MNRAS*, 172, 557  
 Sol H., Pelletier G., Asseo E., 1989, *MNRAS*, 237, 411  
 Tammi J., Duffy P., 2009, *MNRAS*, 393, 1063  
 Tluczykont M., 2011, Proceedings of Texas Symposium 2010. Heidelberg, Germany, Pos(Texas 2010)197  
 Tramacere A., Massaro E., Taylor A. M., 2011, *ApJ*, 739, 66  
 Villata M., Raiteri C. M., 1999, *A&A*, 347, 30  
 Virtanen J. J. P., Vainio R., 2005, *ApJ*, 621, 313  
 Wentzel D. G., 1974, *ARA&A*, 12, 71  
 Yan D., Zhang L., Yuan Q., Fan Z., Zeng H., 2013, *ApJ*, 765, 122  
 Zhdankin V., Uzdensky D. A., Werner G. R., Begelman M. C., 2018, *MNRAS*, 474, 2514  
 Zheng Y. G., Kang S. J., Li J., 2014, *MNRAS*, 442, 3166  
 Zhou Y., Matthaeus W. H., 1990, *J. Geophys. Res.*, 95, 14881

## APPENDIX A: ANALYTICAL SOLUTION OF THE KINETIC EQUATION FOR THE CASE OF SHOCK PERTURBING A STEADY-STATE ELECTRON SPECTRUM

Here we solve the equation (12) describing perturbation of electron population in the VHE emitting zone by a transient shock. Two key parameters of the passing shock govern the evolution of the electron spectrum  $N_{e,FI}(\gamma, t)$  and hence of the MWL emission: the shock acceleration time-scale,  $t_{FI}$ , characterizing the efficiency in acceleration of particles, and  $t_{dur,FI} = t_{cs}$ , which is a duration of the shock acceleration activity in the blob, equal to the passage time of the shock through it (in the frame of the blob). The latter parameter is linked to the time-scale of the flux rise in the light-curve profile of the flare.

### A1 Assumptions and boundary conditions

We assume that the physical parameters of the emitting zone do not change during the passage of the shock. The shock enters the blob at  $t = 0$ .

The initial condition is that the electron spectrum at  $t = 0$  is the steady-state solution  $N_{e,0}(\gamma)$ :

$$N_{e,FI}(\gamma, t = 0) = N_{e,0}(\gamma). \quad (A1)$$

The  $N_{e,0}(\gamma)$  is the asymptotic stationary solution of the kinetic equation with only injection, escape, and cooling terms, and no acceleration process, deduced from the equation (12) (setting shock acceleration term to zero,  $a = 0$ ):

$$\frac{\partial}{\partial \gamma} (b_c \gamma^2 \cdot N_{e,0}(\gamma)) - \frac{N_{e,0}(\gamma)}{t_{esc}} + Q_{inj}(\gamma) = 0. \quad (A2)$$

We neglect the inverse Compton cooling, so the  $b_c$  is constant in time. We also assume  $t_{esc}$  is constant in time and energy-independent:  $t_{esc} \sim R_b/c$ , where  $R_b$  is size of the emitting region. We require a boundary condition such that the electron spectrum tends to zero at the maximal Lorentz factor  $\gamma_{max}$ :  $N_{e,0}(\gamma = \gamma_{max}) = 0$ . With this condition, this equation has the following solution:

$$N_{e,0}(\gamma) = \frac{1}{b_c \gamma^2} \int_{\gamma}^{\gamma_{max}} Q_{inj}(\gamma') \cdot \exp\left(\frac{1/\gamma' - 1/\gamma}{b_c t_{esc}}\right) d\gamma'. \quad (A3)$$

The multiplicative term in this expression and the negative exponent in the exponential describe how the injection effect is respectively damped by the cooling and by the escape. As already discussed, in this paper, we try to not restrict too much the evolution of the particle distribution by artificial boundaries, so we set  $\gamma_{max} \rightarrow \infty$  when evaluating the steady-state electron spectrum given by the

equation (A3). Now let us consider how this spectrum is modified with time when the shock acceleration is acting on this electron population. Let us decompose the electron spectrum in the emitting zone during the passage of the shock into the initial and perturbed parts:

$$N_{e,FI}(\gamma, t) = N_{e,0}(\gamma) + N_{e,p}(\gamma, t). \quad (A4)$$

The initial electron spectrum is the steady-state solution, and the time-dependent perturbed part is the one causing the flux increase. We plug this expression into the kinetic equation (equation 12), which yields:

$$\frac{\partial N_{e,p}(\gamma, t)}{\partial t} = \frac{\partial}{\partial \gamma} (W(\gamma) N_{e,p}(\gamma, t)) - \frac{N_{e,p}(\gamma, t)}{t_{esc}} - \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{t_{FI}} N_{e,0}(\gamma) \right),$$

where  $W(\gamma) = b_c \gamma^2 - \frac{\gamma}{t_{FI}}$ .

This equation describes the time evolution of the perturbed time-dependent addition  $N_{e,p}(\gamma, t)$  to the steady state solution. Here  $t_{FI}$  is assumed constant in time and energy-independent. Plugging the expression for  $N_{e,0}(\gamma)$  (equation A3) to the equation, and evaluating the last term (free term depending only on  $\gamma$ ), we find:

$$\frac{\partial N_{e,p}(\gamma, t)}{\partial t} = \frac{\partial}{\partial \gamma} (W(\gamma) N_{e,p}(\gamma, t)) - \frac{N_{e,p}(\gamma, t)}{t_{esc}} + \mathcal{F}(\gamma) \quad (A5)$$

with  $\mathcal{F}(\gamma) = [1 - \frac{1}{b_c \gamma t_{esc}}] \frac{N_{e,0}(\gamma)}{t_{FI}} + \frac{Q_{inj}(\gamma)}{b_c \gamma t_{FI}}$ .

We obtained the final form of the equation governing how  $N_{e,p}(\gamma, t)$  is evolving with time. The function  $\mathcal{F}(\gamma)$  can be considered as a complex injection function composed of two terms: a scaled steady-state electron spectrum and a scaled injection spectrum. From equations (A1) and (A4) we deduce that the initial condition for  $N_{e,p}(\gamma, t)$ :

$$N_{e,p}(\gamma, t = 0) = 0. \quad (A6)$$

### A2 Solving by characteristics

We use the method of characteristics to solve the equation (A5). We first search for characteristic curves in the  $\gamma-t$  space along which the equation for  $N_{e,p}(\gamma, t)$  becomes an ordinary differential equation. Then we solve this equation along a characteristic curve. Let us rewrite the equation (A5) in the following form (expanding the partial derivative over  $\gamma$ ):

$$\frac{\partial N_{e,p}(\gamma, t)}{\partial t} + (-1) W(\gamma) \frac{\partial N_{e,p}(\gamma, t)}{\partial \gamma} = \mathcal{F}(\gamma) - \frac{N_{e,p}(\gamma, t)}{\tau(\gamma)},$$

where  $\frac{1}{\tau(\gamma)} = \frac{1}{t_{esc}} + \frac{1}{t_{FI}} - 2b_c \gamma$ .

Let us consider a characteristic curve  $(\gamma(t), t)$ . The left-hand side of the equation can be now represented as a full derivative of  $N_{e,p}(\gamma(t), t)$  with respect to time, and also as a directional derivative of  $N_{e,p}(\gamma(t), t)$  in the direction of  $(-W(\gamma), 1)$  in the  $\gamma-t$  plane. By the chain rule, we have:

$$\frac{dN_{e,p}(\gamma(t), t)}{dt} = \frac{\partial N_{e,p}(\gamma(t), t)}{\partial t} + \frac{d\gamma(t)}{dt} \cdot \frac{\partial N_{e,p}(\gamma(t), t)}{\partial \gamma}.$$

We see that along the characteristic curve  $(\gamma(t), t)$  our equation in partial derivatives transforms into an ordinary differential equation:

$$\frac{dN_{e,p}(\gamma(t), t)}{dt} = \mathcal{F}(\gamma) - \frac{N_{e,p}(\gamma(t), t)}{\tau(\gamma)} \quad (A7)$$

$$\frac{d\gamma(t)}{dt} = -W(\gamma). \quad (A8)$$

Let us solve the equation (A8) for the characteristic curve in the  $\gamma$ - $t$  space. We choose an initial point on our characteristic as  $(\xi, 0)$ , so the equation has to satisfy the boundary condition  $\gamma(t=0) = \xi$ . The solution of the equation (A8) with this boundary condition is:

$$\gamma(\xi)(t) = \frac{1}{b_c t_{\text{FI}}(1 - e^{-t/t_{\text{FI}}}) + \frac{1}{\xi} e^{-t/t_{\text{FI}}}}. \quad (\text{A9})$$

This formula defines a characteristic curve in the  $\gamma$ - $t$  space. For given  $\gamma$  and  $t$ , let us find the starting Lorentz factor  $\xi$  of the characteristic that passes through point  $(\gamma, t)$ :

$$\xi = \xi(\gamma, t) = \frac{\gamma e^{-t/t_{\text{FI}}}}{1 - b_c t_{\text{FI}} \gamma (1 - e^{-t/t_{\text{FI}}})}. \quad (\text{A10})$$

Now let us solve the initial value problem (equations A7 and A6). We restrict the  $N_{e,p}(\gamma, t)$  to the characteristic (equation A9), noting  $N_{e,p}(\gamma(\xi)(t), t) \Rightarrow u(t)$  at a given  $\xi$  and solve the differential equation (A7) along the characteristic curve. We have:

$$\frac{du(t)}{dt} + \frac{u(t)}{\tau(\gamma(\xi)(t))} = \mathcal{F}(\gamma(\xi)(t)).$$

This is a simple linear non-homogeneous first-order differential equation, which can be solved with the help of an integrating factor. The equation has the following general solution:

$$u(t) = \frac{1}{\mu(t)} \left[ \int_0^t \mu(t') \mathcal{F}(\gamma(\xi)(t')) dt' + C \right] \quad (\text{A11})$$

with the integrating factor:

$$\mu(t) = e^{\int dt/\tau(\gamma(\xi)(t))}. \quad (\text{A12})$$

From the initial condition equation (A6) which is  $u(t=0) = 0$ , we get the constant of integration  $C = 0$ .

Now let us calculate the  $\mu(t)$  function. First we evaluate the exponent in equation (A12):

$$\begin{aligned} \int \frac{dt}{\tau(\gamma(\xi)(t))} &= \int \left( \frac{1}{t_{\text{esc}}} + \frac{1}{t_{\text{FI}}} \right) dt \\ &\quad - 2b_c \int \frac{1}{b_c t_{\text{FI}} + (1/\xi - b_c t_{\text{FI}}) e^{-t/t_{\text{FI}}}} dt \\ &= \left[ \frac{1}{t_{\text{esc}}} + \frac{1}{t_{\text{FI}}} \right] t - 2 \ln [1/\xi + b_c t_{\text{FI}} (e^{t/t_{\text{FI}}} - 1)]. \end{aligned}$$

The integrating factor  $\mu(t)$  is then:

$$\mu(t) = \frac{e^{(1/t_{\text{esc}} + 1/t_{\text{FI}})t}}{[1/\xi + b_c t_{\text{FI}} (e^{t/t_{\text{FI}}} - 1)]^2}. \quad (\text{A13})$$

### A3 Final solution

The transition from  $u(t)$  back to  $N_{e,p}(\gamma, t)$  is achieved by substitution of  $\xi = \xi(\gamma, t)$  to the expression for  $u(t)$  (equation A11):  $N_{e,p}(\gamma, t) = u(t)|_{\xi = \xi(\gamma, t)}$ .

$$N_{e,p}(\gamma, t) = u(t)|_{\xi = \xi(\gamma, t)} = \int_0^t \left[ \frac{\mu(t')}{\mu(t)} \mathcal{F}(\gamma(\xi)(t')) \right]_{\xi = \xi(\gamma, t)} dt'. \quad (\text{A14})$$

Let us proceed with the substitution  $\xi = \xi(\gamma, t)$  to the components of the integrand.

First we evaluate  $\mu(t)|_{\xi = \xi(\gamma, t)}$ . Substituting the expression for the initial Lorentz factor  $\xi = \xi(\gamma, t)$  from equation (A10), we get:

$$\mu(t)|_{\xi = \xi(\gamma, t)} = \mu(t, \gamma) = \gamma^2 e^{(1/t_{\text{esc}} - 1/t_{\text{FI}})t}. \quad (\text{A15})$$

Next, we compute the form of the  $\mu(t')$  at  $\xi = \xi(\gamma, t)$ , again substituting the expression for  $\xi$  from equation (A10):

$$\begin{aligned} \mu(t')|_{\xi = \xi(\gamma, t)} &= \mu(t', \gamma, t) \\ &= \frac{e^{t'/t_{\text{esc}} + (t' - 2t)/t_{\text{FI}}}}{[1/\gamma + b_c t_{\text{FI}} (e^{(t' - t)/t_{\text{FI}}} - 1)]^2}. \end{aligned} \quad (\text{A16})$$

Then we have to calculate the Lorentz factor  $\gamma(\xi)(t')$  that appears in the function  $\mathcal{F}(\gamma(\xi)(t'))$ . We use equation (A9), (A10) and after simple and obvious transformations we obtain:

$$\gamma(\xi)(t')|_{\xi = \xi(\gamma, t)} = \Gamma(\gamma, t, t') = \frac{\gamma e^{(t' - t)/t_{\text{FI}}}}{1 + \gamma b_c t_{\text{FI}} (e^{(t' - t)/t_{\text{FI}}} - 1)}. \quad (\text{A17})$$

We note that the denominator of  $\mu(t', \gamma, t)$  in equation (A16) multiplied by  $\gamma^2$  is exactly the square of denominator of the  $\Gamma(\gamma, t, t')$ , so for simplicity we express  $\mu(t', \gamma, t)$  via  $\Gamma(\gamma, t, t')$ :

$$\mu(t', \gamma, t) = \Gamma^2(\gamma, t, t') e^{(1/t_{\text{esc}} - 1/t_{\text{FI}})t'}. \quad (\text{A18})$$

Now we evaluate the expression under the integral in equation (A14), using previously derived components, where the substitution was done (equation A15, A18, and A17):

$$\begin{aligned} \left[ \frac{\mu(t')}{\mu(t)} \mathcal{F}(\gamma(\xi)(t')) \right]_{\xi = \xi(\gamma, t)} &= \frac{\Gamma^2(\gamma, t, t')}{\gamma^2} e^{(1/t_{\text{esc}} - 1/t_{\text{FI}})(t' - t)} \\ &\cdot \left[ \frac{N_{e,0}(\Gamma(\gamma, t, t'))}{t_{\text{FI}}} \left( 1 - \frac{1}{b_c t_{\text{esc}} \Gamma(\gamma, t, t')} \right) + \frac{Q_{\text{inj}}(\Gamma(\gamma, t, t'))}{b_c t_{\text{FI}} \Gamma(\gamma, t, t')} \right] \\ &= \frac{\Gamma(\gamma, t, t') e^{(1/t_{\text{esc}} - 1/t_{\text{FI}})(t' - t)}}{b_c t_{\text{FI}} \gamma^2} \\ &\cdot \left[ Q_{\text{inj}}(\Gamma(\gamma, t, t')) + \left( b_c \Gamma(\gamma, t, t') - \frac{1}{t_{\text{esc}}} \right) N_{e,0}(\Gamma(\gamma, t, t')) \right] \end{aligned} \quad (\text{A19})$$

We can now write down the final solution for the total electron spectrum  $N_{e,\text{FI}}(\gamma, t)$  using equation (A4), (A14), and (A19):

$$\begin{aligned} N_{e,\text{FI}}(\gamma, t) &= N_{e,0}(\gamma) + \int_0^t \frac{\Gamma(\gamma, t, t') \cdot \exp[(1/t_{\text{esc}} - 1/t_{\text{FI}})(t' - t)]}{b_c t_{\text{FI}} \gamma^2} \\ &\cdot [Q_{\text{inj}}(\Gamma(\gamma, t, t')) + (b_c \Gamma(\gamma, t, t') - 1/t_{\text{esc}}) \\ &\times N_{e,0}(\Gamma(\gamma, t, t'))] dt'. \end{aligned} \quad (\text{A20})$$

Let us explore the final solution. At the moment when the shock just enters the blob ( $t = 0$ ), the electron spectrum is, as expected, the steady-state solution. Also, when the shock acceleration is extremely weak ( $t_{\text{FI}} \rightarrow \infty$ ), we see that the electron spectrum will remain the steady-state one and not evolve in time, which is in agreement with the expectations (very weak shock will not perturb the electron spectrum).

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.