Electron-beam interaction with emission-line clouds in blazars

Christoph Wendel\textsuperscript{1}, Josefa Becerra González\textsuperscript{2,3}, David Paneque\textsuperscript{4}, and Karl Mannheim\textsuperscript{1}

\textsuperscript{1}Universität Würzburg, 97074 Würzburg, Germany
e-mail: cwendel@astro.uni-wuerzburg.de
\textsuperscript{2}Inst. de Astrofísica de Canarias, 38200 La Laguna, Spain
e-mail: jbecerra@iac.es
\textsuperscript{3}Universidad de La Laguna, Dpto. Astrofísica, 38206 La Laguna, Tenerife, Spain
\textsuperscript{4}Max-Planck-Institut für Physik, 80805 München, Germany
e-mail: dpaneque@mpipi.mpg.de

Received 4 May 2020 / Accepted 7 December 2020

ABSTRACT

Context. An electron-positron beam escaping from the magnetospheric vacuum gap of an accreting black hole interacts with recombination-line photons from surrounding gas clouds. Inverse-Compton scattering and subsequent pair production initiate unsaturated electromagnetic cascades exhibiting a characteristic spectral energy distribution.

Aims. By modelling the interactions of beam electrons (positrons) with hydrogen and helium recombination-line photons, we seek to describe the spectral signature of beam-driven cascades in the broad emission-line region of blazar jets.

Methods. Employing coupled kinetic equations for electrons (positrons) and photons including an escape term, we numerically obtained their steady-state distributions and the escaping photon spectrum.

Results. We find that cascade emission resulting from beam interactions can produce a narrow spectral feature at TeV energies. Indications of such an intermittent feature, which defies an explanation in the standard shock-in-jet scenario, have been found at \( \approx 4\sigma \) confidence level at an energy of \( \approx 3\) TeV in the spectrum of the blazar Mrk 501.

Conclusions. The energetic requirements for explaining the intermittent 3 TeV bump with the beam-interaction model are plausible:

Gap discharges that lead to multi-TeV beam electrons (positrons) carrying \( \approx 0.1\% \) of the Blandford-Znajek luminosity, which interact with recombination-line photons from gas clouds that reprocess \( \approx 1\% \) of the similar accretion luminosity are required.

Key words. galaxies: active – BL Lacertae objects: individual: Markarian 501 – quasars: emission lines – acceleration of particles – scattering – radiation mechanisms: non-thermal

1. Introduction

The launching of astrophysical jets from accreting black holes (BHs) involves fundamental gravitomagnetic processes. The frame-dragging effect on the plasma of the inner accretion flow and its advected magnetic field creates a rotating magnetosphere that can lead to a vacuum voltage drop of \( 10^{20} \) eV for a typical spinning supermassive BH in the centres of galaxies (Levinson & Rieger 2011). It is commonly thought that thermal pair production (PP) in photon-photon collisions provides charge carriers in sufficient numbers to supply the Goldreich-Julian density, which allows electric currents to remove the space charges resulting from the electric potential and to reshape the magnetosphere into a force-free configuration. Any imbalance in the supply of charged particles, however, results in the immediate creation of vacuum (electrostatic) gaps. In vacuum gaps the actual charge density, provided by photon-photon PP (see, however, Petropoulou et al. 2019, for an assessment of the importance of triplet PP in limiting the gap growth), is different from the Goldreich-Julian charge density. Because of this deviation, the force-free condition is not satisfied and the electric field component parallel to the magnetic field lines is non-vanishing and able to accelerate intruding charge carriers to relativistic velocities, extracting energy from the central rotating BH (Blandford & Znajek 1977; Beskin et al. 1992; Hirotani & Okamoto 1998; Levinson 2000).

The interaction of a particle beam with a low-energy radiation field results in electromagnetic cascades. Inverse-Compton (IC) scattering of the primary electrons off low-energy photons produces gamma rays that then undergo PP by interacting with the same photon field leading to a secondary generation of electrons. This sequence repeats until the gamma rays can escape freely (Lovelace et al. 1979; Akharonian et al. 1985; Svensson 1987; Zdziarski 1988; Aharonian & Pshenshnikov 2003).

Cascades have been studied in a variety of astrophysical environments. In outer magnetospheric vacuum gaps of pulsars, curvature radiation is thought to initiate the pair cascade, limiting the gap growth by supplying charged particles (cf. Cheng et al. 1986; Hirotani 2005; Tang et al. 2008). The gamma rays produced by the cascade and leaving the pulsar magnetosphere in a confined beam are made responsible for the observed high-energy (HE) and very high energy (VHE) pulsed emission. In the galactic centre, vacuum gaps are hypothesised to drive cascades that are responsible for PeV cosmic rays and for the VHE gamma-ray point source coincident with Sagittarius A* (cf. e.g., Levinson & Rieger 2011; Katsoulakos et al. 2020).

In active galactic nuclei (AGN) with high accretion rates such as Seyfert galaxies, IC pair cascades in the coronal plasma of the accretion disk are thought to shape their X-ray spectrum up to MeV energies (Svensson 1987; Zdziarski 1988). In AGN with jets such as blazars, unsaturated IC pair cascades can develop from which gamma rays at energies well above...
MeV can escape. The cascades can be initiated by relativistic particles accelerated in vacuum gaps in the BH magnetosphere when the accretion rate is too low to supply the Goldreich-Julian density but their magnetic field is strong enough to launch a jet (Neronov & Aharonian 2007; Neronov et al. 2009; Levinson & Rieger 2011; Hirotani et al. 2017; Katsoulakos & Rieger 2020; Hirotani 2018, for a review). The gaps are thought to be located at the poles of the spinning BH magnetosphere near to the null surface where the Goldreich-Julian charge density is equal to zero (Beskin et al. 1992; Hirotani & Okamoto 1998; Ptitsyna & Neronov 2016; Hirotani et al. 2016; Levinson & Segev 2017; Ford et al. 2018; Chen & Yuan 2020; Kisaka et al. 2020). It was also found that gaps form around the divide of the magnetohydrodynamic (MHD) flow, called the stagnation surface (Vincent & Lebohec 2010; Broderick & Tchekhovskoy 2015; Hirotani & Pu 2016; Aharonian et al. 2017), or near to the inner light surface of the magnetosphere (Cinquin et al. 2020). Time-dependent one-dimensional, general relativistic MHD simulations by Levinson & Segev (2017), Levinson & Cerutti (2018), Chen et al. (2018), and Chen & Yuan (2020) and two-dimensional simulations by Cinquin et al. (2020, for computational reasons, the latter authors had to restrict themselves to unrealistic parameters, however) have recently found inherently non-stationary gap solutions. Levinson & Cerutti (2018) and Kisaka et al. (2020) indicate that the gap activity relaxes to low-amplitude quasi-steady oscillations after an initial spark event, while Chen et al. (2018) and Chen & Yuan (2020) find cyclically enduring gap activity.

Gap-accelerated particles then interact with ambient photon fields, that is, photons from the accretion flow or from the broad line region (BLR), initiating the onset of a cascade (cf. Blandford & Levinson 1995; Broderick & Tchekhovskoy 2015), which gives rise to an additional component of IC up-scattered photons that are emitted along a beam in the same direction as the original electrons were moving. When these cascaded photons can escape from the AGN gamma sphere without being absorbed by PP, and when the line of sight coincides with the beam, the emission could cause an observable imprint on the spectral energy distribution (SED) of the AGN. Intermittency of the particle flux injected from the gap or of the target photon field results in time variability of the emerging radiation and in transient features of the total SED. This could serve as an explanation of the observed minute- and hour-scale variability of HE and VHE emission in blazars (cf. Aharonian et al. 2007; Albert et al. 2007; Foschini et al. 2013; Pitteri et al. 2018; Liao 2018). Similar short-term variability has been found in AGN with jets at inclination angles larger than a few degrees, which supports a similar interpretation (cf. Aleksic et al. 2014; MAGIC Collaboration 2018). Alternative scenarios to explain the short-time variability are the interaction of the jet with the radiation field from a stellar envelope, for instance, or a relativistically propagating emission zone in a relativistically moving jet (cf. Aharonian et al. 2017; Rieger 2019).

In this paper, we examine IC pair cascades in blazars emphasizing the interaction of beam electrons from (short-lived) vacuum gaps with recombination-line photons. Blazars are subdivided into two subclasses, flat-spectrum radio quasars (FSRQs) and BL Lacertae (BL Lac) objects (for a recent review, see e.g., Foschini 2017). These two types were originally distinguished by the presence of broad optical emission lines in the case of FSRQs, with equivalent width $W_\lambda > 0.5$ nm, and the absence of broad optical emission lines for BL Lac objects (Urry & Padovani 1995). However, even in BL Lac objects, in which by definition $|W_\lambda| < 0.5$ nm, optical and ultraviolet recombination lines from the BLR are often present (cf. Stoeck et al. 2011; Becerra González et al. 2020, for Markarian 501 and S4 0954+65, respectively), and might imprint a diagnostically important signature on the gamma-ray spectra emerging from the innermost regions.

This paper is organised as follows: in Sect. 2 we describe the physical model for the beam-photon interaction scenario and its underlying assumptions, including the set of coupled kinetic equations that were employed to numerically evaluate the model. We outline the numerical procedure for solving them to obtain the steady-state particle and photon distributions in Sect. 3. In Sect. 4 we apply this numerical scheme to fit the observational SED of the BL Lac object Markarian 501 (Mrk 501), which exhibited a peculiar spectral feature at $\approx 3$ TeV during a strong flaring activity in 2014 July (MAGIC Collaboration 2020), and discuss the implications of the proposed scenario for the jet formation in Mrk 501. Finally, Sect. 5 summarises our findings.

In the following $m_e$, $c$, $\epsilon$, $\sigma_{\text{Th}}$, $k_B$, and $h$ denote the electron rest mass, the velocity of light, the elementary charge, the Thomson cross section, the Boltzmann constant, and Planck’s constant, respectively. SI units are employed unless noted otherwise.

2. Unsaturated inverse-Compton pair cascade model

The electron beam from an active spark gap region close to the accreting BH can initiate IC pair cascades (Broderick & Tchekhovskoy 2015; Ptitsyna & Neronov 2016; Ford et al. 2018; Levinson & Cerutti 2018; Chen & Yuan 2020). When the beam electrons propagate further along the jet axis, they can encounter the intense photon field of an emission-line cloud as depicted in Fig. 1, and initiate IC pair cascades. Because the threshold energy for PP in photon-photon collisions with the optical and ultraviolet photons characteristic for recombination-line spectra is high, the cascade spectra do not saturate at the electron rest mass energy delivering X-ray spectra, but instead produce spectra peaking at VHE gamma rays. Such cascades are also linear because the reservoir of low-energy target photons is not modified by the cascade in a major way. For their SEDs, it is essential to consider an escape term accounting for the energy-dependent gamma sphere resulting from the pair creation optical depth, which decreases towards the edge of the emission region.

2.1. Unsaturated inverse-Compton pair cascade with escape

For the model of the emission region, we treated electrons and positrons identically and assumed three reservoirs of particle species to be present: firstly, we have relativistic electrons with spectral number density $N(\gamma)$, where $\gamma$ denotes the Lorentz factor of the electron. Secondly, we have highly energetic photons with spectral number density $n_\gamma(x_\gamma)$, where $x_\gamma$ denotes the photon energy in units of the electron rest mass energy $m_e c^2$. In what follows, the term “highly energetic photon” does not only denote photons with energies in the established HE range from 100 MeV to 100 GeV, but more generally photons with energies well above $m_e c^2$, in contrast to low-energy photons with energies well below $m_e c^2$. Thirdly, we have external low-energy photons with spectral number density $n_\nu(x)$ with $x_\nu < 1$.

We assumed that all number densities are time-independent, isotropic, and homogeneous throughout the interaction region, and that interactions take place only inside of it. The interactions were assumed to be dominated by IC scattering and
In a cascade, whose steady-state gamma-ray photon density is $n_\gamma$ and whose escaping flux is $F_{\text{esc}}$. To convert $n_\gamma$ into $F_{\text{esc}}$, cf. Eq. (3).

Because of the interaction of highly energetic photons of energy $x_\gamma$ and low-energy target photons of energy $x_\gamma$, the IC scattering rate is defined as $\gamma'_{\text{IC},\min}(\gamma, x)$ as the minimum possible $\gamma'$ and $x_{\gamma, \text{max}}(\gamma, x)$ as the maximum energy of the up-scattered photon.

Because of the interaction of highly energetic photons of energy $x_\gamma$ with low-energy target photons of energy $x_\gamma$ and spectral number density $n_\gamma(x)$, which result in down-scattered electrons with energy $x_{\gamma, \text{min}}(\gamma, x)$ as defined in Appendix A is the minimum possible $\gamma$ as outlined in Appendix A. Furthermore, we define $\gamma'_{\text{IC},\min}(\gamma, x)$ as the minimum possible $\gamma'$ and $x_{\gamma, \text{max}}(\gamma, x)$ as the maximum energy of the up-scattered photon.

Because of the interaction of highly energetic photons of energy $x_\gamma$ with low-energy target photons of energy $x_\gamma$ and spectral number density $n_\gamma(x)$, which result in down-scattered electrons with energy $x_{\gamma, \text{min}}(\gamma, x)$ as defined in Appendix A is the minimum possible $\gamma$ as outlined in Appendix A. Furthermore, we define $\gamma'_{\text{IC},\min}(\gamma, x)$ as the minimum possible $\gamma'$ and $x_{\gamma, \text{max}}(\gamma, x)$ as the maximum energy of the up-scattered photon.

Because of the interaction of highly energetic photons of energy $x_\gamma$ with low-energy target photons of energy $x_\gamma$ and spectral number density $n_\gamma(x)$, which result in down-scattered electrons with energy $x_{\gamma, \text{min}}(\gamma, x)$ as defined in Appendix A is the minimum possible $\gamma$ as outlined in Appendix A. Furthermore, we define $\gamma'_{\text{IC},\min}(\gamma, x)$ as the minimum possible $\gamma'$ and $x_{\gamma, \text{max}}(\gamma, x)$ as the maximum energy of the up-scattered photon.

Because of the interaction of highly energetic photons of energy $x_\gamma$ with low-energy target photons of energy $x_\gamma$ and spectral number density $n_\gamma(x)$, which result in down-scattered electrons with energy $x_{\gamma, \text{min}}(\gamma, x)$ as defined in Appendix A is the minimum possible $\gamma$ as outlined in Appendix A. Furthermore, we define $\gamma'_{\text{IC},\min}(\gamma, x)$ as the minimum possible $\gamma'$ and $x_{\gamma, \text{max}}(\gamma, x)$ as the maximum energy of the up-scattered photon.
photons is determined from the photon spectral injection rate \( n_{\gamma,i} \), the loss rate (which is the sum of the escape rate with timescale \( T_{\text{esc}} \) and of the pair absorption rate), and the production rate \( n_{\gamma,\text{IC}} \).

Assuming steady state, we obtain, as shown in Appendix A,

\[
\begin{align*}
n_{\gamma}(x) & = 
\frac{\dot{n}_{\gamma,i}(x)}{1 + \int_{0}^{\infty} \gamma \rho(x, y) d\gamma} + \int_{0}^{\infty} \frac{\rho(x, y) d\gamma}{1 + \int_{0}^{\infty} \rho(x, y) d\gamma} \\
& = 
\frac{\dot{n}_{\gamma,i}(x)}{1 + \int_{0}^{\infty} \gamma \rho(x, y) d\gamma} + \int_{0}^{\infty} \gamma \rho(x, y) d\gamma
\end{align*}
\]

(1)

where we have defined two contributions to the photon spectral number density, \( n_{\gamma,i} \), which describes the contribution that arises from the interplay of injection and escape and absorption losses, while \( n_{\gamma,\text{IC}} \) describes the contribution that is achieved from the interplay of production through IC up-scattering and both losses. Exchanging \( y \) and \( \gamma \) in Eq. (1) and plugging this into Eq. (A.8), we can eliminate \( n_{\gamma} \),

\[
N(\gamma) = \frac{N_{\gamma}(\gamma)}{1 + \int_{0}^{\infty} \gamma \rho(x, y) d\gamma} + \int_{0}^{\infty} \gamma \rho(x, y) d\gamma
\]

(2)

3. Numerical solution procedure

In this section, we outline explicit solutions for \( N(\gamma) \) and \( n_{\gamma}(x) \).

The main task is to find solutions for \( N(\gamma) \). This is done by solving Eq. (2) iteratively. The right-hand side of Eq. (2) directly incorporates \( \gamma \), \( T_{\text{esc}} \), \( T_{\text{esc}} \), \( N_{\gamma} \), \( N_{\gamma,i} \), \( C \), \( P \), and the functions in the integration borders, which have been defined in the previous section. \( C \) and \( P \) are known as soon as \( n_{0} \) has been specified. Hence, Eq. (2) can briefly be written as \( N(\gamma) = \mathcal{F}(n_{0}, N_{\gamma}, N_{\gamma,i}, C, P) \), and the functions in the integration borders, which have been defined in the previous section. \( C \) and \( P \) are known as soon as \( n_{0} \) has been specified. Hence, Eq. (2) can briefly be written as \( N(\gamma) = \mathcal{F}(n_{0}, N_{\gamma}, N_{\gamma,i}, C, P) \). As described in Sect. 2.1, the first five arguments of \( \mathcal{F} \) are quantities that describe the physical setting. Hence, they should be understood as input functions and have to be prescribed to specify the physical problem. When this is done, we obtain the equation \( N(\gamma) = \mathcal{F}(N, \gamma) \), where \( \mathcal{F} \) was defined accordingly. Now, we choose a sequence \( \{\gamma_{k}\}_{k=1,\ldots}^{\kappa} \) of \( \kappa \) discrete values \( \gamma_{k} \) along which the kinetic equation is to be solved, in other words, along which \( N(\gamma) \) is to be determined. We write \( N(\gamma_{k}) = N_{\gamma} \) for convenience. As a starting point of the iteration, we guess an initial course \( N_{\text{init}}(\gamma) \) of the function \( N(\gamma) \) and determined the corresponding sequence of values \( (N_{\gamma})_{k=1,\ldots}^{\kappa} \). In Appendix B, we discuss how the converged sequence is determined numerically.

We performed the iteration using the python3 language. To numerically compute the integrals with an infinite integration range, we restricted the integration range we used to a finite range. This is possible when \( n_{\gamma}(x), n_{\gamma,i}(x), \) and \( N_{\gamma}(\gamma) \) are equal to zero everywhere above an upper cut-off, which we denote by \( x_{0} \), \( x_{0} \), and by \( \gamma_{0} \), respectively.

To summarise, the physical setting is specified as soon as the following functions are defined: The low-energy photon spectral number density \( n_{0} \), the highly energetic photon spectral injection rate \( n_{\gamma,i} \), the electron spectral injection rate \( N_{\gamma} \), the highly energetic photon escape time \( T_{\text{esc}} \), and the electron escape time \( T_{\text{esc}} \). By iterating, our code then determined the steady state \( N \). When \( N \) was known, we used Eq. (1) to determine \( n_{\gamma} \), and from this, we obtained the spectral flux density \( F_{\text{esc}} \).

This approach differs from that of Wendel et al. (2017). Therein, although \( T_{\text{esc}} = T_{\text{esc}} = \infty \) (no escape) was used in the kinetic equation, it was assumed that the IC up-scattered photons leave the interaction region from its shell-like boundary. Accordingly, Wendel et al. (2017) directly converted the photon spectral production rate (last term in Eq. (A.7)) into the observed spectral flux density. This appears to be a very crude approach because first it neglects that the injected highly energetic photons can also leave the interaction region (which we take into account by \( n_{\gamma,i} \) in Eq. (1)), and second it assumes that all IC up-scattered photons escape from the boundary of the interaction region and that no IC up-scattered photons escape from its interior. On the one hand, this second assumption contradicts the \( T_{\text{esc}} = T_{\text{esc}} = \infty \) assumption and on the other hand it neglects the energy-dependent attenuation (which we took into account through the division of the photon spectral injection and production rate (numerator of Eq. (1)) by the spectral loss rate (denominator of Eq. (1))).

Based on \( n_{\gamma} \), we determined the spectral flux density \( F_{\text{esc}} \) of highly energetic photons (i.e. the number of photons per unit energy, per unit area, and per unit time) through

\[
F_{\text{esc}}(x) = n_{\gamma}(x) \frac{4\pi c R_{\text{esc}}(x)^{2}}{\Omega_{\gamma} D^{2} m_{e} c^{2}}
\]

(3)

We assumed that the photons leave the interaction region, whose approximate radial size is \( R_{\text{esc}}(x) \), only through a conical beam of opening angle \( \phi \) and solid angle \( \Omega = 4\pi \sin^{2}(\phi/4) \), and that \( D \) is the distance to the observer. The size of the interaction region is linked with the escape time of the photons and electrons. A natural choice for this would be \( R_{\text{esc}}(x) = T_{\text{esc}}(x)c \). The division by \( m_{e}c^{2} \) arises because \( F_{\text{esc}} \) is measured in \( \text{m}^{-2} \text{s}^{-1} \text{J}^{-1} \), while \( n_{i} \) is in units of \( \text{m}^{-3} \).

4. Indication of gap activity in Markarian 501

The numerical procedure described above is used below to explain a peculiar feature in the VHE spectrum of the blazar Mrk 501 during a flaring activity on 2014 July 19.

The broadband emission from blazars, and in particular from the high-synchrotron-peaked BL Lac object Mrk 501, is characterised by a two-bump structure. This emission is usually described within the one-zone synchrotron-self-Compton (SSC) framework (see e.g., Maraschi et al. 1992; Dermer & Giebels 2016). In this model, a blob of relativistic electrons is assumed to move along the jet that points towards the observer. The blob is permeated by a randomly distributed magnetic field, producing synchrotron radiation from the electrons. This radiation is responsible for the low-energy SED bump in blazars. The origin of the HE bump is typically ascribed to the IC scattering of the relativistic electrons with the synchrotron photons. In addition to the SSC scenario, other theoretical models have been proposed to explain the HE and VHE emission in blazars, involving hadronic components and IC pair cascading. Minute-scale variability spectral components mean that the standard one-zone SSC scenario may be insufficient and suggest that an additional sporadic emission component is emitted from the innermost regions of the jet, see Ahnen et al. (2018).

Recently, the MAGIC Collaboration (2020) has presented a two-week flare from Mrk 501, mainly detected in X-rays with
the Swift satellite and in VHE gamma rays observed with the major atmospheric gamma-ray imaging Cherenkov (MAGIC) telescopes. During this flaring state, the source displayed hard X-ray spectra as well as the highest X-ray count rate detected by the Swift X-ray Telescope during its 15 years of operation. In coincidence with the peak of the X-ray count rate, which took place on 2014 July 19 (MJD 56857.98), a narrow feature in the TeV spectrum from the MAGIC telescopes was detected with a significance of $3\sigma$–$4\sigma$. This feature, located around 3 TeV, is inconsistent with the standard functions to describe the VHE spectra from blazars (power law, log-parabola, and log-parabola with exponential cut-off) at more than $3\sigma$ confidence level. The VHE spectrum on 2014 July 19 is better explained assuming a narrow component (a narrow log-parabola or a Gaussian function) in addition to a broad log-parabola. Such a double function fit is preferred with respect to a single log-parabola at $3.2\sigma$–$4.5\sigma$, depending on the a priori assumptions (details can be found in MAGIC Collaboration 2020). This was the first time that a narrow feature (not consistent with the usual smooth spectral functions) was found with a significance higher than $3\sigma$ in the VHE spectrum of Mrk 501, and any TeV blazar in general. Indications for similar (but statistically insignificant) spectral features at $\pm 2$ TeV can be found by inspection of flare spectra reported by the very energetic radiation imaging telescope array system (VERITAS; cf. Fig. 8 of Abdo et al. 2011) and the MAGIC collaboration (cf. Fig. 10 of Ahnen et al. 2017).

Under the assumption that the spectral feature described above is real, three different theoretical scenarios were proposed by the MAGIC Collaboration (2020) to reproduce the narrow spectral feature: (a) a pileup in the electron energy distribution produced by stochastic acceleration, (b) a two-zone SSC model, and (c) a narrow emission from an IC pair cascade scenario induced by electrons accelerated in a magnetospheric vacuum gap. Details for the model describing scenario (c) are provided in this work. In the following, we evaluate the model we fitted to the observational data to infer its implications and consequences.

### 4.1. Specification of the modelled setting

We assumed a vacuum gap near to the poles of the Mrk 501 BH magnetosphere. The potential drop in the gap accelerates seed electrons to ultra-relativistic energies. The seed electrons enter the gap both through PP of gamma rays from the hot advection-dominated accretion flow (ADAF), and possibly through direct leakage from it (Neronov et al. 2009). Seed electrons that have been accelerated in the gap, leave it in the direction of the electric field (i.e. in the direction of the magnetic field), and propagate along a beam that is well collimated to an opening angle $\approx \phi$. During the propagation away from the gap, the seed electrons still encounter the ADAF photon field, and a post-gap cascade (Broderick & Tchekhovskoy 2015; Chen & Yuan 2020) occurs and increases the number of electrons by repeated curvature-radiation and IC emission and PP events. This post-gap cascade ceases when curvature-radiation emission becomes negligible and when IC scattering and PP can no longer be sustained by ADAF photons. Several tens or hundreds of Schwarzschild radii away from the gap, the beam of electrons is assumed to enter a region in which background photons from ionised gas clouds are present. This induces a cascade in this interaction region. This model is depicted in Fig. 1. For this cascade, the model of Sect. 2 was applied. The electron beam from the gap serves as electron injection mechanism in our model. We describe the electron spectral injection rate with the following cut-off Gaussian:

$$N_{\gamma}(\gamma) = \begin{cases} \frac{K_G}{\varsigma \sqrt{2\pi}} \cdot \exp\left(-\frac{(\gamma - \gamma_{\text{mean}})^2}{2\varsigma^2}\right) \text{ if } \gamma_{1,1} \leq \gamma \leq \gamma_{1,0}, \\
0 \text{ otherwise.} \end{cases}$$

(4)

Here, $\gamma_{\text{mean}}$ denotes the mean Lorentz factor of the injected electrons, and $\varsigma$ parametrises the width of the distribution. For the cut-off values we chose $\gamma_{1,1} = \gamma_{\text{mean}} - 3.0\varsigma$ and $\gamma_{1,0} = \gamma_{\text{mean}} + 3.0\varsigma$, satisfying the condition $\gamma \cdot x > 1$ (cf. Sect. 2.1). The normalisation of the Gaussian, which describes the total number of electrons that are injected per unit time and per unit space volume, is called $K_G$.

Furthermore, we assumed that the external low-energy photons of the cascade are represented by emission-line photons from photo-ionised gas clouds in the surroundings of the Mrk 501 BH magnetosphere. Although Mrk 501, as a BL Lac object, does not have a prominent BLR, its BLR has clearly been confirmed observationally, possibly due to outshining by the boosted jet emission, due to a weak accretion rate, or due to the lack of ambient gas (cf. Moles et al. 1987; Stocke et al. 2011). After they are triggered by minor galaxy mergers, interstellar gas clouds, consisting mainly of hydrogen and helium (Wilms et al. 2000), can migrate from the host galaxy into its centre and thus into the central part of Mrk 501. Radiation from the accretion flow or from OB-type stars in the dense galactic centre will inevitably ionise passing clouds. As an effect of recombination, emission-line photons will be abundant in the proximity of the clouds. The external low-energetic photon field can also stem from the envelope of a post-main-sequence star with low metallicity (cf. e.g., Bednarek & Protheroe 1997; Barkov et al. 2010).

In any case, we assumed that the interaction region was both penetrated by the electron beam and pervaded by emission-line photons, and consequently, that an IC pair cascade developed there. It is natural to describe the spectral number density by a sum of Dirac delta distributions,

$$n_0(x) = K_{\text{lines}} \cdot \sum_{i=1}^{4} \frac{K_{\text{line},i}}{x_{0,i}} \cdot \delta_{\text{Dirac}}(x - x_{0,i})$$

(5)

Here, $x_{0,i} = h/(\lambda_{0,i} m_e c)$, and $\lambda_{0,i}$ are the dimensionless energy, and the wavelength of the $i$th line, respectively. $K_{\text{line},i}$ gives the relative flux density contribution of the $i$th line. Precisely, $K_{\text{line},i}$ gives the flux density of the emission line $i$ with respect to the flux density of a hypothetical hydrogen Balmer-$\beta$ line. Accordingly, $K_{\text{line},i}/x_{0,i}$ describes the relative number density contribution of the $i$th line. The parameter $K_{\text{line}}$ is a measure for the photon total number density. We restricted the sum to the three typically most important hydrogen lines plus the most prominent helium line. We compiled these lines and their respective $K_{\text{line},i}$ based on the ultraviolet spectroscopic blazar observations by Pian et al. (2005) and on the synthetic photo-ionisation spectra by Abolmasov & Poutanen (2017). We list them in Table 1.

From an energetic point of view, the emission-line radiation is just a reprocessed fraction of the more luminous ADAF radiation. Nevertheless, we did not include the ADAF radiation in the target photon field. This is justified by the following geometrical argument. The emission of line photons in the ionised cloud is approximately isotropic. The interaction angles of injected electrons and photons are therefore equally distributed; the interaction rate is on average $90^\circ$. This also justifies the usage of a single log-parabola at $\gamma \cdot x > 1$ (cf. Sect. 2.1).
photos, and thus we can neglect the latter. Moreover, except for the reprocessed photons from the ionised cloud, reprocessed ADAF photons might be scarce in the surroundings of the beam due to low ambient gas density in the BL Lac object, which means that they drop out as target photons as well. We have to mention furthermore, that ADAF spectra usually extend to higher energies and \( x \ll 1 \), which was assumed in the derivation of \( C \) and \( P \), is satisfied only very critically.

We did not have injection of highly energetic photons in this setting. Therefore we set the photon spectral injection rate

\[
h_{\gamma,i}(x_i) = 0. \tag{6}
\]

To choose the escape timescales, we determined

\[
T_{\text{phesc}}(x) = T_{\text{esc}}(\gamma) = \frac{R}{c} := T_{\text{esc}}. \tag{7}
\]

In other words, we approximated the interaction volume as a spherical region with radius \( R \). Electrons and photons have equal, and energy independent, that is, constant, escape timescales. The function \( T_{\text{phesc}}(x) \) in Eq. (3) is therefore also set equal to the parameter \( R \).

### 4.2. Modelling results

According to the procedure outlined in Sect. 3, we determined \( N(\gamma) \), \( n_{\gamma}(x) \), and \( F_{\text{esc}}(x,y) \) of the cascade that evolves through interaction of the electrons originally stemming from the gap with the emission-line photons, such that \( F_{\text{esc}}(x) \) was fitted to the narrow SED feature (MAGIC telescopes’ data) and a background SSC emission is fitted to the multi-wavelength SED. We understand \( K_\text{G}, K_\text{lines}, \gamma_\text{mean}, \zeta, R, \) and \( \phi \) as fitting parameters, while we used the luminosity distance \( D = 149.4 \, \text{Mpc} \) (corresponding to redshift \( z = 0.034 \), dimensionless density parameter \( \Omega_m = 0.3 \) of matter, dimensionless density parameter \( \Omega_k = 0.7 \) of dark energy, Hubble constant \( H_0 = 70 \, \text{km s}^{-1} \, \text{Mpc}^{-1} \)) and the line contributions \( K_{\text{line},i} \) given in Table 1. Because the electron beam penetrates into the interaction region from one direction, the kinematics of IC scattering and PP ensure that the highly energetic photons leave the interaction region in this same direction, hence along a beam of opening angle \( \phi \).

Interim results of the numerical procedure are shown in Figs. 3–5 and are discussed now.

The results of our iterative determination of \( N(\gamma) \) are shown in Fig. 3. The initial function of \( N(\gamma) \) was chosen to be

\[
N_{\text{init}}(\gamma) = \frac{2 \cdot \dot{N}(\gamma)}{\int_{\gamma_{\text{min}}}^{\gamma_{\text{max}}} C(\gamma_{i,1}, \gamma') d\gamma' + \int_{\gamma_{\text{min}}}^{\gamma_{\text{init}}(\gamma_{i}-x_i)} C(\gamma_{i,0}, \gamma') d\gamma'}. \tag{8}
\]

cf. Eq. (A.6). This choice is a compromise between a function that can be evaluated fast and is accurate. It is a good approximation for the final \( N(\gamma) \) in the highest energetic regime because the electron spectral production rate is negligible here. For decreasing \( \gamma \), a higher number of iteration steps are necessary because the energy transfer in one IC scattering event decreases. We chose the convergence criterion as follows: One iteration was considered completed when the relative change \( \frac{N(k+1) - N(k)}{N(k)} \) between two successive values was smaller than 0.001 at the points \( \gamma_{k} > 10 \, x_{0,1}^{-1} \) and smaller than 0.001 · \( \frac{(\gamma_{k} x_{0,1})/(10)^{2}}{\gamma_{k} x_{0,1}} \) at the points \( \gamma_{k} \leq 10 \, x_{0,1}^{-1} \). In other words, the demanded accuracy increases quadratically the deeper we iterate into the Thomson regime. There, about \( j_{\text{final}} \approx 100 \) iteration steps were necessary to meet the convergence criterion, while only two or three steps were necessary in the KN regime, cf. Fig. 3. The shallow ridge of \( N(\gamma) \) at \( \gamma \approx 2 \, x_{0,1}^{-1} \) is caused by PP of highly energetic photons with hydrogen Lyman photons (mainly hydrogen Lyman-\( \alpha \) photons), while the very shallow ridge at \( \gamma \approx 0.5 \, x_{0,1}^{-1} \) is caused by PP of highly energetic photons with the helium II Lyman-\( \alpha \) photons. In other words, these ridges in \( N \) are the effect of peaks of the spectral PP rate (fourth summand in Eq. (A.6)). In the Thomson regime, with decreasing \( \gamma \) the course of \( N(\gamma) \) deviates more and more from the \( \gamma^{-2} \) dependence of the saturated cascade case considered by Zdziarski (1988) because particles escape in our scenario. The finite value of the escape time prevents the electrons from penetrating deep into the Thomson regime.

Figure 4 clearly shows that the photon spectral loss rate is dominated by pair absorption in the approximate range \( 3 \, x_{0,1}^{-1} < x_{\gamma} < 800 \, x_{0,1}^{-1} \) corresponding to a photon energy between 19 GeV and 5.1 TeV, and by escape outside this energy range. The maximum of the photon spectral loss rate occurs at \( x_{\gamma} = 12 \, x_{0,1}^{-1} \), corresponding to the maximum of the hydrogen Lyman-\( \alpha \) absorption. The optical depths, which correspond to the respective spectral loss rate contributions, are shown in Fig. 4 on the right-hand side ordinate and were determined by multiplication of the respective spectral loss rate contribution with the escape time \( T_{\text{esc}} \).

We show the photon spectral production rate \( \dot{n}_{\gamma,\text{esc}} \) in Fig. 5. From \( N \), it inherits the peak around 500 \( x_{0,1}^{-1} \) as well as the
deviation from a power law in the Thomson regime due to escape (cf. Zdziarski 1988). In this figure we also show the spectral number density \( n_y \) of highly energetic photons. The contribution \( n_y,IC \) vanishes because of the choice Eq. (6). The contribution \( n_y,IC \) is non-vanishing and ensues as a consequence of IC scattering, pair absorption, and escape. The pronounced trough above \( x_y \approx 4 x_{0,1}^{-1} \) is due to absorption by PP with photons from the hydrogen Lyman-\( \alpha \) line (cf. Fig. 4). Because of the smallness of \( K_{line,2} \) and \( K_{line,3} \), absorption features of the hydrogen Lyman-\( \beta \) line and series are almost invisible. The shallow dip of the spectral number density above \( x_y \approx x_{0,1}^{-1} \) is due to absorption on the helium II Lyman-\( \alpha \) line. Thus, the slight bump at \( x_y \approx x_{0,1}^{-1} \) results from cascaded photons that are immediately below the threshold for PP and hence are not absorbed. The peaky feature around \( 500 x_{0,1}^{-1} \) is more pronounced in \( n_y \) than in \( n_y,IC \) because the spectral loss rate (i.e. the denominator of \( n_y,IC \)) is higher below \( 500 x_{0,1}^{-1} \) than at \( 500 x_{0,1}^{-1} \). If \( n_0 \) were continuous and distributed along a wider range of \( x \) instead of consisting of a few strong lines, the peak around \( 500 x_{0,1}^{-1} \) would be wider and extend to lower energies.

The total energy density of highly energetic photons is obtained by integrating the spectral energy density \( x_y n_y(x_y) \)

\[
\int_{0}^{x_{y,max}(\gamma_0, x_{0,1})} x_y n_y(x_y) \, dx_y \cdot m_e c^2 \approx 4.5 \, \text{J m}^{-3}.
\]

The total energy density of the low-energy photons is \( \sum_{i=1}^{4} K_{line,i} \cdot m_e c^2 \approx 6.5 \, \text{J m}^{-3} \) and thus of similar magnitude. Our cascade is therefore indeed linear (cf. Sect. 2.1).

The cascaded spectral flux density \( F_{casc} \) of highly energetic photons adds to the spectral flux density \( F_{SSC} \) of photons that is produced farther downstream the jet in a common one-zone SSC scenario. Both components contribute to the observed spectral flux density \( F \). We fit both components, \( F_{SSC} \) and \( F_{casc} \), to the multi-wavelength data of Mrk 501 of the night of MJD 56857.98 as presented by the MAGIC Collaboration (2020). As shown in Fig. 10 of MAGIC Collaboration (2020) and in Fig. 6 in the present paper, we can describe the narrow SED feature without the need of extreme SSC parameters. The parameters found by the fit are given for the cascaded emission component in Table 2 of the present paper and for the SSC emission component in Table 6 and Fig. 7 in MAGIC Collaboration (2020). The emission from a cascade initiated by electrons originating from a vacuum gap can thus explain the observed narrow peak-like feature.

4.3. Physical inferences

In this subsection we discuss the results of modelling Mrk 501. We assess the findings in a broader context.

4.3.1. Number and size of emission-line clouds

We compared the assumed hydrogen Lyman-\( \alpha \) photon abundance with the hydrogen Lyman-\( \alpha \) luminosity of \( \int L_{\alpha,YSO,\, obs} = 5.2 \times 10^{33} \) W that was detected from Mrk 501 by Stocke et al. (2011). The total number density of all low-energy photons is given by \( K_{lines,0} \cdot \sum_{i=1}^{4} K_{line,i} / x_{0,i} \approx 3.2 \times 10^{18} \, \text{m}^{-3} \). This number density describes the photon field of the ionised cloud assumed to be interacting with the electron beam at the gap. The total number density of hydrogen Lyman-\( \alpha \) photons is determined by \( K_{lines} \cdot K_{line,4} / x_{0,4} \approx 2.6 \times 10^{18} \, \text{m}^{-3} \). We further assumed that
there are $N_{\text{cl}}$ clouds of approximate radial size $R_{\text{cl}}$ in the BLR. The hydrogen Lyman-α luminosity $L_{\text{Ly,α,cl}}$ of all these clouds is responsible for the observed hydrogen Lyman-α luminosity. We can set

$$L_{\text{Ly,α,obs}} = L_{\text{Ly,α,cl}} = N_{\text{cl}} \cdot 4\pi R_{\text{cl}}^2 \frac{K_{\text{line,cl}}}{x_0} \frac{c \times 0.4 \times m_e e^2}{c^2}.$$ (10)$$

As BL Lac objects are considered to be evolved objects running short of cold gas supplies (Liodakis 2018; Foschini 2017), we can assume that $N_{\text{cl}}$ is quite small. On the other hand, to obtain a Gaussian line profile for the emission lines resulting from the Doppler broadening, the number of clouds must not be too small. Then, solving Eq. (10) for the radial size yields an estimate of $R_{\text{cl}}$. Choosing $N_{\text{cl}} = 10$ yields $R_{\text{cl}} \approx 1.8 \times 10^{11} \text{m}$, which roughly agrees with our value used for the radial size of the interaction region $R$ (cf. Table 2). This size of the cloud corresponds to the size of a red giant, which is suggestive of cloud formation by photospheric abrasion through beam interaction. In terms of Schwarzschild radii $r_S$ for a central BH of mass $M = 10^8 \text{M}_\odot$, it is $R_{\text{cl}} = 0.061 r_S$. From the modelling we infer that in Mrk 501 only few major gas clouds are present in the BLR. The encounter of a major cloud with electrons from the gap-beam must therefore be a rare event. Presuming that the gas clouds revolve around the BH on unperturbed Keplerian orbits, an encounter of a cloud with the beam of electrons will at best repeat with the Keplerian orbital period of the cloud. These conclusions are in line with the fact that the peak-like SED feature has not been found elsewhere in Mrk 501 data.

4.3.2. Reprocessing of the accretion flow luminosity

The total line luminosity of these $N_{\text{cl}}$ clouds of radius $R_{\text{cl}}$ is determined as

$$L_{\text{lines, tot}} = N_{\text{cl}} \cdot 4\pi R_{\text{cl}}^2 \frac{K_{\text{line,cl}}}{x_0} \sum_{i=1}^{4} K_{\text{line},i} c \times m_e e^2.$$ (11)$$

For the above chosen $N_{\text{cl}} = 10$ and resulting $R_{\text{cl}} = 1.8 \times 10^{11} \text{m}$, it is $L_{\text{lines, tot}} \approx 7.8 \times 10^{33} \text{W}$, a factor 5 lower than the upper limit of the Mrk 501 BLR luminosity found by Celotti et al. (1997). We recall that this finding was obtained from our modelling and from the observed hydrogen Lyman-α luminosity. In canonical BLR models, the BLR luminosity is assumed to be about 10% of the accretion flow luminosity (Ghisellini et al. 2017). As Mrk 501 is a gas-depleted object, its BLR reprocesses even less of the incident accretion flow radiation. Thus, the reprocessing fraction $\xi = L_{\text{lines, tot}} / L_{\text{ADAF, tot}}$ might be lower than 10%, say 1%. Here, $L_{\text{ADAF, tot}}$ is the total luminosity of the ADAF. To compare the emission-line luminosity with the accretion flow luminosity, we applied the synthetic ADAF spectrum computed by Mahadevan (1997). There, the spectral luminosity of an ADAF as a function of frequency was parametrised by the viscosity parameter $\alpha$, the pressure ratio $\beta$, the inner boundary $R_{\text{in}}$ of the hot accretion flow, its outer boundary $R_{\text{out}}$, the central BH mass $M$, the electron temperature $T_e$, and by the dimensionless mass accretion rate $\dot{m}$, which is defined as the ratio of the mass accretion rate to the Eddington accretion rate. We integrated numerically, and for the purpose of cross checking also analytically, along the photon energy $\epsilon$ over the spectral luminosity $L_{\text{ADAF}}(\epsilon)$ to obtain $L_{\text{ADAF, tot}}$. We integrated essentially over the superposition of the cyclosynchrotron emission, its Comptonised part, and of the bremsstrahlung contribution. We used the canonical choices $\alpha = 0.3$, $\beta = 0.5$, $R_{\text{in}} = 3 r_S$, and $R_{\text{out}} = 1000 r_S$ as well as $M = 10^8 \text{M}_\odot$. Furthermore, we chose the energy $\epsilon_{\text{ADAF, 0}} = 300 k_B T_e$ as an upper integration border because the spectrum essentially vanishes above this energy. We let $\dot{m}$ and $T_e$ be free parameters. The chosen requirement $L_{\text{lines, tot}} \approx 0.01 \cdot L_{\text{ADAF, tot}}(\dot{m}, T_e)$ can be met with the exemplary $(\dot{m}, T_e)$ pairs listed in row 1 of Table 3. This shows that the present scenario fits the general picture of accretion flows as illuminators of BLRs. Moreover, it gives evidence that Mrk 501 is indeed fed by a low-rate, hot ADAF. When the reprocessing fraction $\xi$ is increased from 1% to the canonical 10%, the values of $T_e$ decrease by about 25% (cf. Table 3).

4.3.3. Constraints on the in-gap and post-gap multiplication

The parameter $K_G$ expresses the number density of electrons that intrude into the interaction region per time interval. In our model, these electrons enter the interaction region after their number was increased in the gap and during the post-gap cascade. The origin of the electron beam along which the multiplication takes place is the vacuum gap (cf. Fig. 1), where in turn the origin of the seed electrons is PP through self-interaction of ADAF radiation. In other words, photons

---

**Table 2.** Fitting parameters that were used to describe the narrow SED feature by emission from an IC pair cascade.

| Quantity          | Used value                  |
|-------------------|-----------------------------|
| $\phi$            | 1.8°                        |
| $R$               | $3.0 \times 10^{11} \text{m}$ |
| $K_G$             | $3.3 \times 10^2 \text{m}^{-3}$ |
| $K_{\text{lines}}$ | $9.7 \times 10^2 \text{m}^{-3}$ |
| $\gamma_{\text{mean}}$ | $3.4 \times 10^2 \text{eV}/(m_e c^2)$ |
| $\epsilon$        | $0.23 \gamma_{\text{mean}}$ |

---

Fig. 6. High-energy bump of the SED of Mrk 501 from 2014 July 19 (MJD 56857.98). The dashed grey line depicts the IC bump of a one-zone SSC model (contribution $F_{\text{SSC}}$). The dot-dashed red line depicts the emission of the cascade (contribution $F_{\text{esc}}$, cf. green line in Fig. 5). The sum $F$ of both contributions is depicted by the solid black line. The spectral data from the MAGIC telescopes are shown as red circles, while measurements and upper limits by the Fermi Large Area Telescope are drawn by black and yellow triangles. The details on the data analysis and the SSC modelling can be found in MAGIC Collaboration (2020).
that are emitted by the ADAF can collide in the BH magnetosphere and pair-produce seed electrons. We can determine the number density of ADAF photons capable of PP with the help of the ADAF model by Mahadevan (1997). The spectral number density of ADAF photons is given by $n_{\text{ADAF}}(\epsilon) = \lambda_{\text{ADAF}}/(c \cdot \epsilon \cdot \epsilon_{\text{ADAF}})$, where $\lambda_{\text{ADAF}} = 2 \pi R_{\text{out}}^2 + 2 \pi R_{\text{in}}^2 + 2 \pi \cdot (R_{\text{out}} \sqrt{1.25 R_{\text{out}}^2 - R_{\text{in}}^2} - R_{\text{in}} \sqrt{1.25 R_{\text{in}}^2})$ is a rough estimate of the surface area of the ADAF. We determined the total surface area as a sum of the outer cylindrical lateral area and the inner cylindrical lateral area, and twice the outer conical lateral area was subtracted by the inner conical lateral area. Additionally, we assumed the height of the ADAF to be $r$ at radius $r$.

As mentioned above, the ADAF spectrum is non-vanishing up to the energy $\epsilon_{\text{ADAF},0}$. Because PP is a threshold process, the lowest possible ADAF photons capable of PP have the energy $\epsilon_{\text{ADAF,PP,th}} = m_e c^2 / \epsilon_{\text{ADAF},0}$. Consequently, the total number density of ADAF photons capable of PP is obtained by

$$n_{\text{PP,1}} \approx \int_{\epsilon_{\text{ADAF,PP,th}}}^{\epsilon_{\text{ADAF,0}}} n_{\text{ADAF}}(\epsilon) \, d\epsilon.$$  \hspace{1cm} (12)

As a comparison, we determined the total number density of pair-producing MeV photons based on the approximation

$$n_{\text{PP,2}} \approx 1.4 \times 10^{17} \frac{m^2}{M} 10^9 M_{\odot}^{-1} \text{m}^{-3}.$$  \hspace{1cm} (13)

from Levinson & Rieger (2011), which is motivated by findings for the bremsstrahlung luminosity of an ADAF by Narayan & Yi (1995). For the ADAF parameters we used above, the assumed $\xi = 1\%$, and the constrained pairs of $n$ and $T_e$, we give the correspondingly values of $n_{\text{PP,1}}$ and $n_{\text{PP,2}}$ in the second and third row of Table 3. Except for the last pair, we have an order-of-magnitude agreement. In contrast, for the canonical $\xi = 10\%$ and the corresponding lower $T_e$, we obtain values for $n_{\text{PP,1}}$ that are about two orders of magnitude lower (cf. Table 3) than in the 1% reprocessing fraction case.

From the number density of pair-producing ADAF photons, we estimated the materialisation rate in the gap. With $0.2 \sigma_{\text{th}}$ as an approximation for the PP cross section, the number density of pairs that are produced in the gap region per unit time is $K_{\text{PP,gap}} \approx 0.2 \sigma_{\text{th}} n_{\text{PP,1}} c$. We show the values of $K_{\text{PP,gap}}$ for the $(n, T_e)$ pairs in the fourth row of Table 3. All four estimates in the $\xi = 1\%$ case are about six orders of magnitude smaller than our determined $K_G$ and all four estimates in the $\xi = 10\%$ case are about ten orders smaller. This does not necessarily cast doubt on our model.

These pair-produced seed electrons are accelerated in the gap and multiplied in the gap and during the propagation to the interaction with the emission-line photons. The multiplication is mediated by the emission of synchro-curvature radiation in the magnetospheric magnetic field, by the emission of IC up-scattered radiation in the ADAF photon field, and by the subsequent materialisation of this radiation through PP. In net effect, this multiplication increases the number of electrons by the number $N$ and also lowers the energy per electron by $N$. Because we neither included the magnetic field and the emission of synchro-curvature radiation nor the exact spatial distribution of the ADAF photon field into our model, a detailed analysis of the in-gap and post-gap multiplication is beyond the scope of this work. We can, however, estimate from comparison of $K_{\text{PP,gg}}$ with $K_G$ that it has to be $N \approx 10^6$ in the 1% reprocessing fraction case and $N \approx 10^{10}$ in the 10% case. The finding $N \approx 10^6$ is three orders of magnitude larger than the multiplication found by Broderick & Tchekhovskoy (2015). These authors, however, determined the multiplication only in the post-gap cascade. Considering that up to $10^5$ curvature photons can be emitted by one electron in the initial spark event of the gap (Levinson & Cerutti 2018), $N \approx 10^6$ after the post-gap cascade is plausible and does not inhibit gap stability. Furthermore, to assess the magnitude of $N$, we have to consider the magnetic field structure (especially the divergence of the magnetic field), which can dilute the number density of the electron bunch.

The maximum potential drop in the gap is connected with the maximum Lorentz factor of the electrons. Knowing from our modelling (cf. Eq. (4) and Table 2) that the maximum energy per electron after the multiplication is $\gamma_{\text{max}} m_c^2 = (\gamma_{\text{mean}} + 3\xi) m_c^2 = 5.7 \times 10^{12} \text{eV}$, we can estimate that the maximum energy that one original seed electron has attained in the gap voltage is $\approx N \cdot \gamma_{\text{max}} m_c^2$. In the $\xi = 1\%$ case, this gives $\approx 5.7 \times 10^{18} \text{eV}$, which agrees with estimated (Levinson & Rieger 2011) and simulated (Hirotani & Pu 2016) values for the potential drop in a vacuum gap. In the $\xi = 10\%$ case, this gives $\approx 5.7 \times 10^{22} \text{eV}$, which seems hard to reach even for extreme magnetospheric gap conditions. Thus, the canonical choice $\xi = 10\%$ for the reprocessing fraction does not fit a consistent picture of Mrk 501. Levinson & Segev (2017), Levinson & Cerutti (2018), Chen & Yuan (2020), Crinquand et al. (2020), and Kisaka et al. (2020) have recently cast doubt on the prior attitude of vacuum gaps being steady-state phenomena. Instead, they might be of intermittent or cyclic character. In particular, Levinson & Cerutti (2018) found that charge-starved gaps, after ignition by injection of gamma rays or electrons abruptly discharge in

---

**Table 3. Four combinations of physical quantities connecting ADAF properties to the materialisation rate in the vacuum gap.**

| $(n, T_e)$ | $n_{\text{PP,1}} / \text{m}^{-3}$ | $n_{\text{PP,2}} / \text{m}^{-3}$ | $K_{\text{PP,gg}} / \text{s}^{-1} \text{m}^{-3}$ |
|------------|---------------------------------|---------------------------------|---------------------------------|
| $(1.0 \times 10^{-4}, 1.7 \times 10^{10} \text{K})$ | $2.8 \times 10^9$ | $1.4 \times 10^9$ | $0.031$ |
| $(2.0 \times 10^{-4}, 1.4 \times 10^{10} \text{K})$ | $4.6 \times 10^9$ | $5.6 \times 10^9$ | $0.084$ |
| $(4.0 \times 10^{-4}, 1.1 \times 10^{10} \text{K})$ | $4.7 \times 10^9$ | $2.2 \times 10^{10}$ | $0.088$ |
| $(8.0 \times 10^{-4}, 8.4 \times 10^{9} \text{K})$ | $5.5 \times 10^9$ | $9.0 \times 10^{10}$ | $0.058$ |
| $(1.0 \times 10^{-4}, 1.3 \times 10^{10} \text{K})$ | $6.7 \times 10^9$ | $5.8 \times 10^9$ | $1.2 \times 10^{-5}$ |
| $(2.0 \times 10^{-4}, 1.0 \times 10^{10} \text{K})$ | $3.8 \times 10^9$ | $1.9 \times 10^{-5}$ | $0.058$ |

**Notes.** The first row shows pairs of values for the dimensionless accretion rate and for the electron temperature of an ADAF that are permitted by reconciliation of the ADAF luminosity with the luminosity of the line emitting clouds. The corresponding values of the total number density of pair producing ADAF photons according to the model by Mahadevan (1997), of the total number density of pair producing ADAF photons according to the estimation by Levinson & Rieger (2011), and of the materialisation rate of pair-produced electrons in the gap region are shown in the second, third, and fourth line, respectively. In each field, the upper item was yielded with the assumption of $\xi = 1\%$ and the lower item was yielded with $\xi = 10\%$. As $n_{\text{PP,2}}$ is independent of $T_e$, it has the same value for both cases.
that cross the particle beam. Numerically solving the coupled kinetic equations corresponding to linear IC pair cascades with escape terms, we obtained the steady-state solution of the SED of the photons escaping from the interaction zone. Compared to previous studies of cascades in AGN, our approach considered three main ramifications: Firstly, the inclusion of an escape term corresponds to a source that is finitely extended and not necessarily optically thick. In contrast to the work of Zdziarski (1988), which dealt with saturated and hence optically thick cascades, our scenario can therefore be applied to astrophysical objects that are optically thin to the emission of HE gamma rays. This is only a first step towards the self-consistent treatment of a spherically symmetric emission region with a radially varying emission and absorption coefficients. For the SSC mechanism without pair cascades, such a treatment was developed by Gould (1979). In our framework, the photon density reacts back upon itself through the electron density. We therefore did not treat the emission coefficient as prescribed. Secondly, Wendel et al. (2017) determined the photon spectral number density $n_\gamma$ essentially as the accumulation of all the IC up-scattered photons. In contrast, we determined $n_\gamma$ by considering both IC up-scattering and losses through escape and pair absorption (cf. Eq. (1)). Thirdly, we used an efficient numerical iteration scheme (cf. Sect. 3) that saves computational cost by iterating from values $y_\gamma$ with high $k$ to lower $\tilde{k}$ and so using the optimal course of $N(y)$ in each iteration step.

Applying the model to the blazar Mrk 501, we showed that the recently observed peculiar narrow spectral feature at 3 TeV can readily be explained when rather plausible physical parameters for the beam and recombination-line clouds are adopted. In the scenario supported by this observation, photons from a weakly accreting hot ADAF materialise as pairs in the central BH magnetosphere. These seed electrons are accelerated in a vacuum gap and are multiplied by post-gap cascades while travelling away from the central region and carry about 0.1% of the Blandford-Znajek luminosity. The ADAF also irradiates and ionises ambient gas clouds. In agreement with measurements of the hydrogen Lyman-α luminosity in Mrk 501, these clouds reprocess about 1% of the original ADAF luminosity as emission-line photons. They act as a target colliding with the electron beam. The resulting cascade photons are emitted in the direction of the primary beam electrons, shaping the intermittent narrow bump at about 3 TeV in the SED. In combination with the SSC spectrum produced by particles presumably accelerated at shock waves further downstream, the model explains the broadband SED of Mrk 501 at MJD 56857.98, assuming no further pair attenuation within the host galaxy to occur. Predictions about the duty cycle of this feature are difficult considering the intermittency of gap formation and randomness of nearby cloud passages.

References

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2011, ApJ, 727, 129
Abolmasov, P., & Poutanen, J. 2017, MNRAS, 464, 152
Agaronyan, F. A., Atoyan, A. M., & Nagapetyan, A. M. 1983, Astrophysics, 19, 187

A115, page 10 of 14
Appendix A: Derivation of the kinetic equations

As an attachment to Sect. 2.2, here we give a precise derivation of the kinetic equations. To quantify the change rates of the spectral number densities of the photons and of the electrons, it is necessary to introduce the spectral interaction rates of IC scattering and of PP. We do this in line with Zdziarski (1988).

We assume an IC scattering event of an incident electron with original energy $\gamma \gg 1$ off a photon with energy $x \ll 1$ that results in a down-scattered electron with final energy $\gamma'$ and a highly energetic photon with energy $x'$. Energy conservation yields $\gamma \approx \gamma + x = \gamma' + x'$. From IC scattering kinematics, it follows

$$\gamma x > E_x \quad (A.1)$$

with the abbreviation $E_x = (\gamma/\gamma' - 1)/4$ (cf. Zdziarski 1988).

This ensures that the photon energy has a maximum possible value $x_{\gamma, \max}(x, \gamma) := (4\gamma x)/(1 + 4\gamma x)$. This function is $< \gamma$ for all realistic values of $\gamma$, which reflects that the electron cannot transfer all its energy to the photon. Moreover, we remark that $x_{\gamma, \max}(x, \gamma)$ increases with increasing $\gamma$ as well as with increasing $x$. Hence, if $N$ has an upper cut-off at $\gamma_i$ and $n_0$ has an upper cut-off at $\gamma_0$, then the distribution of highly energetic photons vanishes above $x_{\gamma, \max}(\gamma_i, 0)$ and is non-vanishing below this.

The spectral IC scattering interaction rate $C$ for such events on the field of low-energy photons with spectral number density $n_0$ is given by Zdziarski (1988) as an approximation of the exact one found by Jones (1968),

$$C(\gamma, \gamma') = \int_{x_{\gamma, \max}(\gamma, x, \gamma)}^{x_{\gamma, \max}(\gamma, \gamma, x)} n_0(x) \frac{3\sigma_{\text{Th}} c}{4\gamma^2 x} \frac{E_x}{\gamma x} \left[ r + (2 - r) \frac{E_x}{\gamma x} - 2 \frac{E_x}{\gamma x} \ln \frac{\gamma x}{E_x} \right] \, dx. \quad (A.2)$$

Here it is $r = (\gamma/\gamma' + \gamma/\gamma')/2$ and the second argument in the lower integration border is due to Eq. (A.1). $N(\gamma)C(\gamma, \gamma')$ is the number of IC scattering events with original energy $\gamma$ and final energy $\gamma'$ per unit time, per unit space volume, per unit original energy, and per unit final energy.

Substituting $\gamma'$ with $\gamma - x'$ in $C$, we realise that $C(\gamma, \gamma - x')$ describes the probability for an IC scattering event with original electron energy $\gamma$ to produce a photon with energy $x'$, per unit time, per unit $x'$-interval. $N(\gamma)x(y, \gamma - x')$ stands for the number of IC scattering events with original electron energy $\gamma$ that result in a photon with energy $x'$, per unit time, per unit space volume, per unit $y'$-interval, and per unit $x'$-interval.

Now, we consider a collision of a highly energetic photon of energy $x_0$ with a photon of energy $x$, resulting in the PP of an electron of energy $\gamma$ and a positron with energy $\gamma + x - \gamma \approx x_0$. Based on Agaronyan et al. (1983), Zdziarski (1988) gave the corresponding kinematic relationship

$$x_0 x > E_x > 1 \quad (A.3)$$

where $E_x = x_0^2/(4\gamma(x_0 - \gamma))$ as well as the spectral PP interaction rate $P$, again for a photon field $n_0$,

$$P(x_0, \gamma) = \int_{x_{\gamma, \max}(\gamma, x, \gamma)}^{x_{\gamma, \max}(\gamma, \gamma, x)} n_0(x) \frac{3\sigma_{\text{Th}} c}{4\gamma^2 x} \frac{E_x}{\gamma x} \left[ r + (2 - r) \frac{E_x}{\gamma x} + 2 \left( \frac{E_x}{x_0 x} \right)^2 + 2 \frac{E_x}{x_0 x} \ln \frac{x_0 x}{E_x} \right] \, dx. \quad (A.4)$$

With this, we can quantify the number of IC scattering events with final energy $\gamma'$ and any permitted original energy $\gamma$ per unit time, per unit space volume, and per unit final energy by $\int_{y_{\text{min}}(\gamma)}^{\gamma} N(\gamma)C(\gamma, \gamma') \, dy$. The upper integration border is $y_{\text{min}}(\gamma)$ in contrast to the one erroneously used by Zdziarski (1988) in the second term on the right-hand side of his Eq. (1) as well as in the first integral in A21.

Fourth, an electron loses energy by becoming IC down-scattered to a lower energy $\gamma' < \gamma$, resulting in a sink at $\gamma$. We can infer from Eq. (A.1) that a final electron energy $\gamma'$ cannot be reached from any original electron energy $\gamma$. For $\gamma' \geq 1/(4x)$ (scattering into the KN regime), the original electron energy is indeed unbounded. However, for $\gamma' < 1/(4x)$ (scattering into the Thomson regime), the original electron energy has to be lower than $\gamma'(1 - 4\gamma')$, which increases with increasing $\gamma$. Again, if the low-energy photon spectral number density is extended with an upper cut-off at $x_0$, then the upper boundary of $\gamma'$ is expressed by $\gamma'(1 - 4\gamma')$. In what follows we define

$$\gamma_{\text{IC, max}}(\gamma, y) := \begin{cases} \gamma/(1 - 4\gamma') & \text{for } \gamma' < 1/(4x), \\ \infty & \text{for } \gamma' \geq 1/(4x). \end{cases} \quad (A.5)$$

Fifth, PP of an electron with energy $\gamma$ is a source at this energy. However, PP on a photon with energy $x$ that produces an electron with energy $\gamma$ is possible only if the photon energy $x_0x$ exceeds the PP threshold $x_{\gamma, \text{PP, th}}(x, \gamma) := \gamma/(1 - (1/(4x))^2)$. This threshold can be inferred from Eq. (A.3). $x_{\gamma, \text{PP, th}}$ adopts finite and positive values only for $\gamma > 1/(4x)$, expressing that electrons cannot be pair-produced below $1/(4x)$. In this range, $x_{\gamma, \text{PP, th}}$ has a minimum at $\gamma = 1/(2x)$. This minimum corresponds to
With this in mind, the number of PP events with incident photon energy $\gamma$ per unit time, per unit space volume, and per unit electron energy is accordingly expressed by $\int_{x_{\gamma,PP}(\gamma,x_0)}^{\gamma_{x,PP}} n_\gamma(x_\gamma) P(x_\gamma, \gamma) \, dx_\gamma$. As $P$ is normalised to the production of only one electron, this term has to be multiplied by 2 in the kinetic equation.

The electron kinetic equation is obtained by setting the rate of change $N(\gamma)$ of the electrons’ spectral number density equal to the sum of all source terms subtracted by all sinks,

$$N(\gamma) = \dot{N}(\gamma) - N(\gamma) \left( \frac{1}{t_{esc}(\gamma)} + \int_0^\gamma C(\gamma', \gamma) \, dy' \right)$$

$$+ \int_{\gamma_{IC,th}}^{\gamma_{x,PP}} N(\gamma') C(\gamma', \gamma) \, dy'$$

$$+ 2 \cdot \int_{x_{\gamma,PP}(\gamma,x_0)}^{\gamma_{x,PP}} n_\gamma(x_\gamma) P(x_\gamma, \gamma) \, dx_\gamma.$$  \hfill (A.6)

In what follows the first term of this equation is called electron spectral injection rate, the part within round brackets is called electron spectral loss rate (which is the sum of the spectral escape rate and of the spectral pair absorption rate), and the last term is called the electron escape rate and of the spectral IC down-scattering rate to lower energies, and the sum of the third and fourth term are called electron spectral production rate (which is the sum of the spectral IC down-scattering rate to higher energies and of the spectral PP rate). We note that the dimension of the electron spectral injection rate and of the electron spectral production rate is $[\gamma^{-1} \cdot t_{esc}^{-1}]$, respectively, and that of the electron spectral loss rate is $[\gamma^{-1} \cdot t_{esc}^{-1}]$.

Analogously, we identify all source and sink terms that determine the rate of change $\dot{n}_\gamma(x_\gamma)$ of the highly energetic photon spectral number density: First, the injection term $\dot{n}_\gamma(x_\gamma)$ has to be included. Second, the spectral loss rate due to escape from the interaction volume is described by $n_\gamma(x_\gamma) / t_{esc}(x_\gamma)$.

Third, PP is a sink of the highly energetic photon distribution at the energy $x_\gamma$ of the incident photons. For given $x_\gamma$ and $x_\gamma$, the energies $\gamma$ that can be reached by the created electrons, are limited due to the interaction kinematics Eq. (A.3). The electron energy $\gamma$ that can be obtained obey $\gamma_{PP, min} < \gamma < \gamma_{PP, max}$ with the limits $\gamma_{PP, min}(x_\gamma, x_\gamma) := x_\gamma(1 - (1 - 1/(x_\gamma, x_\gamma))^{1/2})/2$ and $\gamma_{PP, max}(x_\gamma, x_\gamma) := x_\gamma(1 + (1 + 1/(x_\gamma, x_\gamma))^{1/2})/2$. With increasing $x_\gamma$, the limit $\gamma_{PP, min}$ decreases and $\gamma_{PP, max}$ increases. Consequently, if $n_\gamma$ is extended and non-vanishing only below and at an upper cut-off $x_\gamma$, then the electron energies that are kinematically allowed satisfy $\gamma_{PP, min}(x_\gamma, x_\gamma) < \gamma < \gamma_{PP, max}(x_\gamma, x_\gamma)$. With this in mind, the number of PP events with incident photon energy $x_\gamma$, and any possible electron energy per unit time, per unit space volume, and per unit x_\gamma-interval can be written as

$$\int_{x_{\gamma,PP}(\gamma,x_0)}^{\gamma_{x,PP}} n_\gamma(x_\gamma) P(x_\gamma, \gamma) \, dx_\gamma.$$  \hfill (A.7)

In what follows the first term of this equation is called photon spectral injection rate, the part within round brackets is called photon spectral loss rate (which is the sum of the spectral escape rate and of the spectral pair absorption rate), and the last term is called photon spectral production rate. We remark that the dimension of the photon spectral injection rate and of the photon spectral production rate is $[x_\gamma^{-1} \cdot t_{esc}^{-1} \cdot t_{esc}^{-1}]$, and that of the photon spectral loss rate is $[x_\gamma^{-1} \cdot t_{esc}^{-1} \cdot t_{esc}^{-1}]$. We assume the spectral number densities to be in steady state. Thus, the respective rates of change vanish in Eqs. (A.6) and (A.7). Then, we can solve the respective equation for the respective spectral number density in the second term on the right-hand side of the respective equation. This yields

$$N(\gamma) = \dot{N}(\gamma) + \int_{x_{IC,th}(\gamma,x_0)}^{\gamma_{IC,th}} N(\gamma') C(\gamma', \gamma) \, dy'$$

$$+ 2 \cdot \int_{x_{IC,th}(\gamma,x_0)}^{\gamma_{x,PP}} n_\gamma(x_\gamma) P(x_\gamma, \gamma) \, dx_\gamma.$$  \hfill (A.8)

and Eq. (1).

### Appendix B: Discussion of numerical solution procedure

As an extension to Sect. 3, in this appendix we concisely explain how explicit converging solutions for $N(\gamma)$ are found from iteratively solving $N(\gamma) = \mathcal{F}(N, \gamma)$ along the sequence of points $(\gamma_j)_{j=1, \ldots, n}$ after prescribing the initialisation $N_{init}(\gamma)$. We could iterate the following way: We could determine $\mathcal{F}$ at all the points $\gamma_k$ based on the sequence $(N_{init})_{init}$. Then, we would assign the attained values to a new sequence $(N_j)$.

In short, the initialisation step of the iteration would be described by $N_0 = \mathcal{F}((N_{init})_{init}, \gamma_k)$. The same technique would be pursued in the subsequent iteration steps. In each iteration step, we would compute a complete new sequence $(N_j)$ (with integer $j$ running along $\gamma_k = 0, 1, \ldots, \gamma_{final}$ and denoting the number of the iteration step, or equivalently, the number of the sequence $(N_j)$). In other words, in the $j$th iteration step, we would determine the $x$ values $N_{x,j} = \mathcal{F}((N_{x,j-1}, \gamma_k)$ based on the sequence...
$(N_{k})_{1}$ of the previous iteration. Iterating would have to proceed until all points $N_{k,j}$ converge with a value called $N_{j,\text{final}}$ (and hence the function $N(\gamma)$ converges), which is always the case according to the Banach fixed-point theorem.

Wendel et al. (2017) used this iteration scheme. However, this scheme is not favoured because it is computationally inefficient. This is shown as follows. By inspection of Eq. (2) we realise that a computation of $F$ at a given point $\gamma_k$ evaluates (and thus needs to know) $N$ only at points $\gamma_k \geq \gamma_k$. This observation is obvious in the case of the second term in the numerator of Eq. (2) because the integration range stretches from $\gamma$ upwards. In the case of the third term, we have to consider the values of $\gamma$ in the range $\gamma_k$, thus needs to know (we consider the computation of a certain value $N_{\text{final}}$; the final value being $\gamma_k$). This procedure is advantageous because during the iteration at a point $\gamma_k$, the final converged sequence $(N_{k})_{k=0,1,...,\text{final}}$ of values is already known and can be used, which reduces inaccuracies at $\gamma_k$.

In the range $\gamma > 1/(4\kappa_0)$ pairs can be produced. This is reflected in the third term in the numerator of Eq. (2) being non-zero. This makes the computation of $F$ expensive in this regime. At $\gamma \leq 1/(4\kappa_0)$, PP does not happen and the third term vanishes, which speeds up the computations. On the other side, more iteration steps are necessary for lower values of $\gamma$ due to the decreasing energy transfer in one IC scattering event. In the deep Thomson regime (i.e. for $\gamma \ll 1/(4\kappa_0)$) Eq. (A.6) in the steady state can be expressed as a continuity equation and could be solved by a differential equation solver (e.g., a Runge-Kutta method, cf. Zdziarski 1988). However, as we are mainly interested in the course of $N$ in the KN regime, we used the iteration procedure over the complete range.