Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** 11587–11626 (2024) https://doi.org/10.1093/mnras/stad3863
Advance Access publication 2023 December 20

Kinetic simulations and gamma-ray signatures of Klein–Nishina relativistic magnetic reconnection

J. Mehlhaff [1,2★] G. Werner [2] B. Cerutti [1] D. Uzdensky [2] and M. Begelman [3,4]

[1] *Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France*
[2] *Center for Integrated Plasma Studies, Physics Department, University of Colorado, 390 UCB, Boulder, CO 80309, USA*
[3] *JILA, University of Colorado and National Institute of Standards and Technology, 440 UCB, Boulder, CO 80309, USA*
[4] *Department of Astrophysical and Planetary Sciences, University of Colorado, 391 UCB, Boulder, CO 80309, USA*

Accepted 2023 December 12. Received 2023 December 11; in original form 2023 June 29

ABSTRACT

Black hole and neutron star environments often comprise collisionless plasmas immersed in strong magnetic fields and intense baths of low-frequency radiation. In such conditions, relativistic magnetic reconnection can tap the magnetic field energy, accelerating high-energy particles that rapidly cool by inverse Compton (IC) scattering the dense photon background. At the highest particle energies reached in bright gamma-ray sources, IC scattering can stray into the Klein–Nishina regime. Here, the Comptonized photons exceed pair-production threshold with the radiation background and may thus return their energy to the reconnecting plasma as fresh electron–positron pairs. To reliably characterize observable signatures of such Klein–Nishina reconnection, in this work, we present first-principles particle-in-cell simulations of pair-plasma relativistic reconnection coupled to Klein–Nishina and pair-production physics. The simulations show substantial differences between the observable signatures of Klein–Nishina reconnection and reconnection coupled only to low-energy Thomson IC cooling (without pair production). The latter regime exhibits strong harder-when-brighter behaviour; the former involves a stable spectral shape independent of overall brightness. This spectral stability is reminiscent of flat-spectrum radio quasar (FSRQ) GeV high states, furnishing evidence that Klein–Nishina radiative physics operates in FSRQs. The simulated Klein–Nishina reconnection pair yield spans from low to order-unity and follows an exponential scaling law in a single governing parameter. Pushing this parameter beyond its range studied here might give way to a copious pair-creation regime. Besides FSRQs, we discuss potential applications to accreting black hole X-ray binaries, the M87* magnetosphere, and gamma-ray binaries.

Key words: acceleration of particles – magnetic reconnection – radiation mechanisms: general – relativistic processes – gamma-rays: general.

1 INTRODUCTION

The gamma-ray sky is studded with relativistic compact objects – neutron stars and black holes (of which the most numerous observed varieties are, respectively, pulsars and blazars: Wakely & Horan 2008; Abdollahi et al. 2020). These systems – and connected phenomena including winds, jets, and accretion discs – frequently host collisionless highly magnetized plasmas, with magnetic energy density exceeding not just the pressure (i.e. small plasma beta) but also the rest-mass energy density of the charge-carrying particles. In such plasmas, relativistic magnetic reconnection (Blackman & Field 1994; Lyutikov & Uzdensky 2003; Lyubarsky 2005; Watanabe & Yokoyama 2006) efficiently siphons off the excess magnetic field energy, using it to accelerate relativistic particles and drive relativistic collective motion. The energized particles are then revealed by the light that they shine toward Earth, including in the gamma-ray band. Relativistic magnetic reconnection is, hence, an important candidate mechanism for powering high-energy phenomena linked to the most compact objects in the Universe.

In some astrophysical situations, there is a well-defined time-scale separation between abrupt reconnection-powered particle acceleration and much slower radiative losses. In this radiatively inefficient regime, observable emission traces particle energization that has occurred in the past. This limit is seldom realized, however, in the plasma environments of compact objects, where intense magnetic and radiation fields lead to rapid synchrotron and inverse Compton (IC) cooling. Then, the problem can no longer be cleanly factorized into a sudden acceleration step followed by a more prolonged cooling stage. Instead, radiative cooling couples in real time to reconnection, tracing active (as opposed to past) particle acceleration, and feeding back on the reconnection process: a qualitatively distinct radiative reconnection regime (Uzdensky 2011, 2016; Mehlhaff et al. 2021).

In the low-energy, optically thin limit where the synchrotron and IC photons freely escape the system – what we might call *classical radiative reconnection* – radiative losses, while dynamically important, do not change the fundamental flow of energy from the non-radiative case. Magnetic fields serve as the main energy source

★ E-mail: john.mehlhaff@univ-grenoble-alpes.fr

for particle acceleration, while the primary energy sink is the emission mechanisms that efficiently and permanently remove liberated magnetic energy from the system. However, for the gamma-ray-bright relativistic compact objects, the photons emitted at the highest energies are above threshold for various quantum electrodynamical (QED) pair production channels. This fundamentally alters the pathways available to the energy in radiative reconnection, allowing radiation to not just carry energy away from the reconnection site, but also to redistribute it in real time in the form of freshly produced electron–positron pairs. Such *QED radiative reconnection* is thus distinguished from merely classical radiative reconnection in its capacity to alter the plasma material composition and in the key role played by photons as a dynamically important particle species (Uzdensky 2011, 2016; Uzdensky et al. 2019).

To interpret observations of systems where QED reconnection may occur, modelling efforts must employ a self-consistent kinetic plasma description. Such a description is already needed to model collisionless relativistic magnetic reconnection in the non-radiative and classical radiative regimes, because it captures the critical microphysics governing the reconnection rate (i.e. the rate of magnetic energy dissipation) as well as the production of non-thermal particle energy distributions and correspondingly non-thermal emission spectra. In the case of QED reconnection, a kinetic paradigm is even more imperative. The QED cross-sections depend sensitively on the energies of both the emitting particles and the pair-producing photons, placing an even greater importance on capturing energization self-consistently.

All of the necessary kinetic physics can be incorporated by augmenting *ab initio* particle-in-cell (PIC) simulations (Hockney & Eastwood 1988; Birdsall & Langdon 1991) with QED physics. The small number of PIC reconnection studies that have done this have focused on a select few QED interactions. For example, Hakobyan, Philippov & Spitkovsky (2019) present a regime, expected in pulsar magnetospheres, where particles suffer strong synchrotron cooling, and the resulting synchrotron photons, with an emission spectrum peaking at $\gtrsim$ MeV energies, collide with one another, leading to copious pair production in the reconnection inflow region (see also Hakobyan, Ripperda & Philippov 2023b). The same emission and pair-production channels are coupled to reconnection (with an additional IC post-processing step) by Xie et al. (2023) in the context of magnetar magnetospheres. Reconnection in magnetar magnetospheres is also studied by Schoeffler et al. (2019, 2023). However, those authors consider pair production not between colliding synchrotron photons, but from the absorption of single synchrotron photons by the intense magnetar magnetic field. They observe the radiative cooling to remove particle pressure support from the reconnection layer, leading to strong plasma compression there. This locally amplifies synchrotron emission and pair creation, with the end result that pair production in this regime is concentrated not in the plasma fuelling reconnection, but in the heart of the reconnection layer itself. Lastly, Crinquand et al. (2021, 2022) present global models of reconnection in black hole magnetospheres, where the primary radiation mechanism is IC scattering of low-energy (soft) background photons originating from a larger-scale accretion flow. Pair production then occurs when Comptonized photons collide with the soft parent population from which they were first scattered, supplying the plasma to a luminous equatorial reconnection current sheet in the black hole magnetosphere.

The examples above illustrate two general points. First, while QED reconnection is of general high-energy astrophysical interest, the relevant QED interactions depend on the system under study. Second, specializing to certain QED interactions over others not only decides the applicable astrophysical sources, but can also lead to divergent qualitative dynamics. These remarks underscore the need to understand QED reconnection – in all its astrophysical diversity – as a fundamental physics problem in order to identify its observable signatures in the high-energy Universe.

In this work, we present PIC simulations run using the ZELTRON code (Cerutti et al. 2013; Cerutti & Werner 2019, plus needed auxiliary developments detailed here) of a QED reconnection regime thus far unexplored from first principles as an isolated physics problem. We consider a relativistic magnetic reconnection layer immersed in such an intense bath of soft background radiation that IC scattering strays far into the QED limit. This contrasts the (classical radiative) low-energy, Thomson IC reconnection regime previously studied numerically by Werner, Philippov & Uzdensky (2019), Mehlhaff et al. (2020), Sironi & Beloborodov (2020), and Sridhar, Sironi & Beloborodov (2021, 2023) in two respects. First, we account for quantized gamma-ray emission from the highest energy particles radiating in the Klein–Nishina IC limit (Jones 1968; Blumenthal & Gould 1970). Second, because many of the emitted photons lie above pair-production threshold with the soft background (Mehlhaff et al. 2021), we calculate pair production between the few (low number density) scattered gamma-rays with energies $\epsilon_{\rm scat} \gg m_e c^2$, and the abundant (high number density) soft seed photons with energies $\epsilon_{\rm ph} \ll m_e c^2$. We refer to this realization of QED reconnection as *Klein–Nishina radiative reconnection* (sometimes just *Klein–Nishina reconnection*), omitting explicit reference to pair production since efficient Klein–Nishina IC emission implies pair production in a reconnection context (Mehlhaff et al. 2021).

The QED interactions studied here are the same as those treated by Crinquand et al. (2021, 2022). Here, however, we take a complementary approach, stripping away the global morphology and studying reconnection as a local problem. Computationally, this enables us to concentrate resources toward enhancing the separation among the radiative and plasma microscales. Physically, it permits us to remain more agnostic to the host system, focusing instead on the intrinsic reconnection properties that may be generally applicable.

Indeed, Klein–Nishina reconnection may be realized in a range of astrophysical systems connected to relativistic compact objects, including: flat-spectrum radio quasars (FSRQs), where reconnection occurring in a relativistic jet launched from an active galactic nucleus (AGN) is likely externally illuminated by large-scale circumnuclear structures (cf. Giannios, Uzdensky & Begelman 2009; Nalewajko et al. 2011, 2012; Giannios 2013; Petropoulou, Giannios & Sironi 2016; Sironi, Giannios & Petropoulou 2016; Werner et al. 2018; Christie et al. 2019, 2020; Giannios & Uzdensky 2019; Mehlhaff et al. 2020, 2021; Ortuño-Macías & Nalewajko 2020); the high/soft states of accreting black hole X-ray binaries (BHXRBs), where reconnection in a highly magnetized collisionless coronal region is illuminated by an underlying geometrically thin, optically thick accretion disc (cf. Galeev, Rosner & Vaiana 1979; Di Matteo 1998; Goodman & Uzdensky 2008; Uzdensky & Goodman 2008; Hoshino & Lyubarsky 2012; Uzdensky 2016; Beloborodov 2017; Werne et al. 2019; Sironi & Beloborodov 2020; Mehlhaff et al. 2021; Sridhar et al. 2021, 2023; El Mellah et al. 2022); the magnetospheres of supermassive black holes, particularly the one at the centre of the M87 galaxy, M87*, wherein reconnection may be bathed in photons from a large-scale radiatively inefficient accretion flow (cf. Giannios, Uzdensky & Begelman 2010; Ball et al. 2016; Li, Yuan & Wang 2017; Ripperda, Bacchini & Philippov 2020; Crinquand et al. 2021, 2022; El Mellah et al. 2022; Ripperda et al. 2022; Scepi, Dexter & Begelman 2022; Chen, Uzdensky & Dexter 2023; El Mellah, Cerutti & Crinquand 2023; Galishnikova et al. 2023;

Hakobyan et al. 2023b); and gamma-ray binaries, where a plausible scenario involves a pulsar in tight orbit around a bright type O or Be star, which illuminates reconnection occurring near the pulsar in its magnetosphere and striped wind (i.e. before the pulsar wind shocks with the stellar wind from the companion; cf. Dubus 2006; Cerutti, Dubus & Henri 2008; Dubus 2013; Cerutti & Philippov 2017; Dubus et al. 2017; Philippov & Spitkovsky 2018; Cerutti, Philippov & Dubus 2020). The link between Klein–Nishina reconnection and each of these object classes is, in fact, a major result of this work (Section 6), as further discussed below.

While this study is primarily numerical, analytic, and semi-analytic modelling are also vital for understanding Klein–Nishina reconnection and QED reconnection more broadly. Such theoretical approaches can make targeted, physically motivated arguments for how results from non-QED reconnection may generalize to the QED case (e.g. Beloborodov 2017; Mehlhaff et al. 2021; Chen et al. 2023; Hakobyan et al. 2023b), even if they cannot treat all of the kinetic physics at play from first principles. This furnishes a useful interpretive framework for *ab initio* simulations. However, the reverse is also true: phenomenological models, which sometimes have the advantage of enhanced physical clarity, can themselves be refined from the findings of simulations. This study illustrates both directions of this paradigm. Throughout the text, we make frequent reference to our earlier work, Mehlhaff et al. (2021), which analytically considers the setup simulated here. As will be seen, that study (besides laying much of the theoretical foundation for this article) serves both as an interpretive lens for our simulations and as a set of hypotheses that the numerical experiments can check.

We structure this article as follows. In Section 2, we detail the QED algorithmic developments that enable our PIC simulations. Then, in Section 3, we describe our simulation setup in detail. In Section 4, we present the results of our simulations, comparing and contrasting Klein–Nishina radiative reconnection to two control cases: one of non-radiative reconnection and one of classical radiative reconnection subject to efficient Thomson IC losses. Section 5 then provides a second results section. However, there, instead of delving into a detailed analysis of a few simulations with different radiative physics, we conduct parameter scans with all of the QED physics turned on, characterizing the pair yield of reconnection – a single number computed per simulation – as a function of its main controlling parameters. In Section 6, we survey observations of the four main application systems targeted by this work – FSRQs, BHXRBs, the M87* magnetosphere, and gamma-ray binaries – discussing connections to our simulation results. We conclude with a complete summary of our findings in Section 7. In the remaining part of this Introduction, we preview the three principal astrophysical results of this study.

The first concerns the correlated spectral and temporal signatures of Klein–Nishina reconnection (Section 4.5). As in the non-radiative and classical (Thomson IC) radiative regimes, Klein–Nishina reconnection powers efficient non-thermal particle acceleration (NTPA) and, hence, non-thermal radiative emission. While the time-averaged observable spectrum is similar to that of Thomson IC reconnection, the relationship between the shape of the output spectral energy density and its luminosity are very different. In Thomson IC reconnection, these two are tightly correlated, with a shallower observed spectrum coinciding with a higher luminosity (i.e. 'harder-when-brighter'). In Klein–Nishina reconnection, this correlation is broken: the spectrum exhibits a constant shape irrespective of overall brightness. These results, potentially observable during gamma-ray flares, represent an important distinguishing property of Klein–Nishina reconnection and illustrate the importance of collecting temporally resolved observed spectra.

Our second main astrophysical result concerns the electron–positron pair yield of Klein–Nishina reconnection (Section 5). Using simulations, we derive an empirical formula for the Klein–Nishina reconnection pair yield in terms of a single control parameter. While, in the regime probed by this study, reconnection generally produces, at most, order-unity new pairs per processed pair, our derived scaling law, together with physical arguments for its extrapolation, point to a potential regime where Klein–Nishina reconnection may be a copious pair source. However, even in the case of order-unity pair yield, Klein–Nishina reconnection can still convert an initially electron-ion plasma into a strongly mixed electron-ion-positron plasma. This reconnection regime is thus a potentially important *in situ* antimatter source in astrophysics.

Our final main astrophysical result is a detailed survey of the four main object classes – FSRQs, BHXRBs, the M87* magnetosphere, and gamma-ray binaries – where Klein–Nishina reconnection may occur (Section 6). We find potentially strong observational connections to GeV observations of FSRQs, where observed spectral stability during flares is reminiscent of the anticipated spectral-temporal signatures of Klein–Nishina reconnection. We also sketch a roadmap of the theoretical and instrumental developments necessary to link Klein–Nishina reconnection modelling more rigorously to observations in the other systems. We comment on the potential effect of the Klein–Nishina reconnection pair yield on the global operation and observable aspects of each examined object type.

2 SIMULATION ARCHITECTURE

We here detail the QED capabilities that we added to the radiative electromagnetic PIC code ZELTRON (Cerutti et al. 2013; Cerutti & Werner 2019) to enable the simulations presented in this article. Readers wishing to skip these technical details altogether may proceed directly to Section 3; those interested in only a general description may prefer covering just Section 2.1 before moving on. Excellent additional references on QED methods in PIC simulations can be found in the literature documenting other PIC codes, including: TRISTAN v2 (e.g. Hakobyan et al. 2019, 2023a), OSIRIS (e.g. Fonseca et al. 2002; Del Gaudio et al. 2020), SMILEI (Derouillat et al. 2018), WARPX (Fedeli et al. 2022), and GRZELTRON (e.g. Levinson & Cerutti 2018; Crinquand et al. 2020). With the exception of the null collision method (see Section 2.2.3), the algorithms employed by us are the same as those used in GRZELTRON (cf. the supporting information of Crinquand et al. 2020). However, to our knowledge, many of the specifics – especially the optimization techniques of Section 2.2.3 – have not yet been described to the same level of detail as follows.

2.1 The QED PIC method

To provide some context for the modifications we have made to the ZELTRON code, we first review salient general features of electromagnetic PIC codes (though see, e.g., Hockney & Eastwood 1988; Birdsall & Langdon 1991 for a more thorough treatment), discussing how they may be extended to include QED effects. Fig. 1 provides a graphic summary of this discussion.

The electromagnetic PIC (hereafter, simply 'PIC') technique is a computational method for simulating first-principles kinetic plasma physics. PIC simulations are kinetic in that they self-consistently describe the full phase space (position + momentum) plasma distribution function (in contrast to, e.g., fluid plasma frameworks,

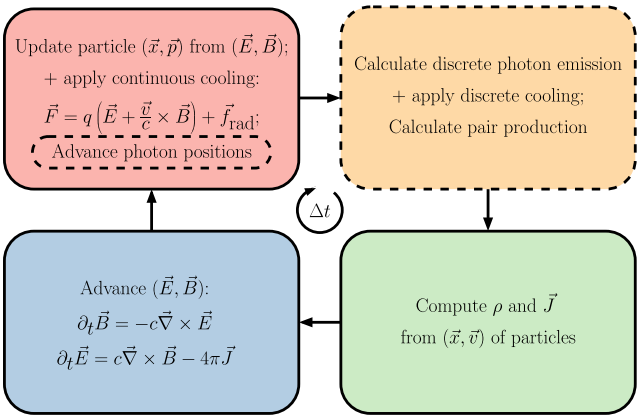

Figure 1. Standard PIC loop augmented with steps (dashed outline) to model QED effects. At each PIC time-step, the electromagnetic field – plus an optional continuous radiative drag force to model non-QED cooling – are used to update the particles' positions and momenta according to the Lorentz force law (top left). Photons are also ballistically propagated at the speed of light in this step. Then (top right), QED cross-sections are used to compute Monte Carlo photon emission and pair production. Each time a Monte Carlo photon is emitted, its momentum is self-consistently subtracted from that of its radiating particle. Freshly emitted photons and produced pairs are added to the simulation. In the next step (bottom right), the particle positions and velocities are used to calculate the electromagnetic charge and current densities. These are used in the final step (bottom left) to update the electromagnetic fields via the time-dependent Maxwell's equations.

which track bulk quantities – such as spatial density and local mean velocity – in real space only). The PIC technique is, furthermore, a first-principles method because it evolves physical equations (the Maxwell–Vlasov system) requiring minimal approximations. Owing to these properties, PIC simulations can probe detailed microscopic plasma physical effects, while furnishing vital, self-consistent astrophysical observables such as light curves and spectra.

The PIC method represents the simulated plasma as a large number of discrete charged particles coupled to electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{B}$, respectively. The $\boldsymbol{E}$ and $\boldsymbol{B}$ fields are tracked on a spatial grid, while the particles' positions, $\boldsymbol{x}$, and momenta, $\boldsymbol{p}$, can vary continuously. At each time-step, the gridded field values are interpolated to the positions of the particles, allowing their momenta to be evolved via the Lorentz force law,

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=q\left(\boldsymbol{E}+\frac{\boldsymbol{v}}{c}\times\boldsymbol{B}\right) \tag{1}$$

(Fig. 1, red panel; $\boldsymbol{v}$ is the 3-velocity for a particle of momentum $\boldsymbol{p}$). Then, to capture the response of the fields to the particles, the particles' positions and momenta are used to calculate the bulk charge, ρ, and current, $\boldsymbol{J}$, densities on the computational field grid (Fig. 1, green panel). This allows $\boldsymbol{E}$ and $\boldsymbol{B}$ to be advanced via the time-dependent Maxwell's equations,

$$\begin{aligned}\partial_t\boldsymbol{B}&=-c\boldsymbol{\nabla}\times\boldsymbol{E}\quad\text{and}\\ \partial_t\boldsymbol{E}&=c\boldsymbol{\nabla}\times\boldsymbol{B}-4\pi\boldsymbol{J}\end{aligned} \tag{2}$$

(Fig. 1, blue panel). It is worth noting that equations (1) and (2) imply a collisionless PIC method, since the particles only interact collectively through the long-range fields $\boldsymbol{E}$ and $\boldsymbol{B}$ rather than individually through short-range two- or few-body encounters.

By itself, the procedure described so far does not model the high-energy radiation (with photon frequencies unresolved by the simulation time-step) that is often important in the environments of relativistic compact objects, as in this work. The simplest way that such radiative effects can be incorporated is by adding a cooling radiative drag term, $\boldsymbol{f}_{\rm rad}$, to equation (1) such that it becomes (Fig. 1, red panel)

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=q\left(\boldsymbol{E}+\frac{\boldsymbol{v}}{c}\times\boldsymbol{B}\right)+\boldsymbol{f}_{\rm rad}\,. \tag{3}$$

Treating radiative cooling as a continuous drag force is suitable when the radiating particles do not lose a significant fraction of their energy to any single photon emission event. However, when particles begin to emit photons at energies that rival their own, radiation becomes inherently discrete, and a more general approach is necessary. As an additional concern, when the emitted photons are above the threshold energy for one or more pair-production processes, their propagation and absorption must be handled self-consistently.

The needed additional QED operations can be mostly consolidated into one extra step beyond the standard PIC loop (Fig. 1, orange panel). Here, the particle positions and velocities are used – perhaps in conjunction with the electromagnetic fields – to evaluate QED cross-sections for the photon emission and pair-production processes of interest, yielding probabilities for these events to occur. A subset of the possible events are triggered by comparing their probabilities with randomly drawn numbers – a Monte Carlo procedure – and the resulting photons and particles are added to the simulation. Besides this additional QED Monte Carlo step (Fig. 1, orange panel), the positions of photons also need to be evolved. This is typically done alongside the particles (Fig. 1, red panel) and is much easier by comparison because photons follow straight lines (in flat space–time, as in this work). We term the PIC method, expanded to include these extra operations as diagrammed in Fig. 1, the *QED PIC* method.

To enable this work, we have generalized the Cartesian version of the PIC code ZELTRON to include QED physics as sketched here. The emission and pair-production processes – relevant to the orange panel of Fig. 1 – are: IC scattering (including the high-energy Klein–Nishina limit) of a soft seed photon background by ultrarelativistic particles; and pair production when the high-energy Comptonized photons are absorbed by the soft background. In the following discussion, we provide a sketch of the algorithm used to model these processes.

2.2 QED effects implemented in ZELTRON

2.2.1 Monte Carlo inverse Compton emission procedure

We begin by describing algorithmic details of the IC emission mechanism. In this section, we present a conceptually simple but computationally expensive implementation. In Section 2.2.3, we discuss the modifications we made to the simpler procedure of this section for the sake of optimization.

Our simulations are immersed in a homogeneous, static, isotropic, and monochromatic photon radiation bath of energy density (per unit energy interval)

$$u(\epsilon)=U_{\rm ph}\delta(\epsilon-\epsilon_{\rm ph})\,. \tag{4}$$

These background photons are not tracked by the simulation; their energy density is prescribed by hand. Electrons and positrons, through the IC process, can upscatter these background photons. If an upscattered photon attains final energy above the (gamma-ray) threshold for pair production with the background radiation bath, then it is promoted to a tracked photon whose position is evolved explicitly by the simulation.

For a particle with ultrarelativistic Lorentz factor $\gamma \gg 1$ traversing the radiation field (4), the number of photons scattered per unit time is given as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = c\sigma_{\mathrm{T}} n_{\mathrm{ph}}\, g_{\mathrm{KN}}\left(\gamma/\gamma_{\mathrm{KN}}\right) , \tag{5}$$

where σ_{T} is the Thomson cross-section, $n_{\mathrm{ph}} \equiv U_{\mathrm{ph}}/\epsilon_{\mathrm{ph}}$ is the background photon number density,

$$\gamma_{\mathrm{KN}} \equiv m_e c^2/4\epsilon_{\mathrm{ph}} \tag{6}$$

is the critical Lorentz factor above which IC scattering transitions to the Klein–Nishina regime, and $g_{\mathrm{KN}}(q) \leq 1$ is the dimensionless function (calculated in appendix A of Mehlhaff et al. 2021)

$$g_{\mathrm{KN}}(q) = \frac{3}{2q^2}\left[\left(q+9+\frac{8}{q}\right)\ln(1+q) - \frac{1}{1+q}\left(\frac{q^2}{2}+9q+8\right) + 4\,\mathrm{Li}_2(-q)\right] . \tag{7}$$

Here, $\mathrm{Li}_2(q)$ is the dilogarithm. As needed to recover the Thomson regime, $g_{\mathrm{KN}}(\gamma/\gamma_{\mathrm{KN}})$ tends to unity in the low-energy limit, $\gamma \ll \gamma_{\mathrm{KN}}$. Over a simulation time-step Δt, equation (5) yields a probability

$$p_{\mathrm{emit}} = \Delta t\, \mathrm{d}N/\mathrm{d}t \tag{8}$$

for the particle to scatter a photon. We operate in the regime $p_0 \equiv c\sigma_{\mathrm{T}} n_{\mathrm{ph}} \Delta t \ll 1$, implying $p_{\mathrm{emit}} = p_0 g_{\mathrm{KN}}(q) \leq p_0 \ll 1$.

Computationally, the Monte Carlo photon emission process can be accomplished, for each particle at each time-step, by evaluating the probability (8) and drawing a random number to determine whether the emission event occurs. Then, for the subset of particles that actually emit photons, the scattered photon energy ϵ_{scat} must be assigned. This demands drawing a second random number $\mathcal{R}$ and inverting the probability distribution over ϵ_{scat}: that is, finding ϵ_{scat} such that

$$\mathcal{R} = \frac{\int_0^{r(\epsilon_{\mathrm{scat}})} K(r', q)\mathrm{d}r'}{\int_0^1 K(r', q)\mathrm{d}r'} . \tag{9}$$

Here, $q \equiv \gamma/\gamma_{\mathrm{KN}}$ as above, $r(\epsilon_{\mathrm{scat}})$ is a proxy for the photon energy defined by

$$r(\epsilon_{\mathrm{scat}}) \equiv \frac{\epsilon_{\mathrm{scat}}/\gamma m_e c^2}{q\left(1-\epsilon_{\mathrm{scat}}/\gamma m_e c^2\right)} , \tag{10}$$

and $K(r, q)$ is the single-particle Klein–Nishina scattering kernel as reported by Jones (1968) and Blumenthal & Gould (1970):

$$K(r,q) = \frac{3}{(1+qr)^2}\left[2r\ln r + (1+2r)(1-r) + \frac{1}{2}\frac{(qr)^2}{1+qr}(1-r)\right] . \tag{11}$$

That is, the number of photons scattered per unit time by a particle with $\gamma = q\gamma_{\mathrm{KN}}$ to final proxy photon energy between r and $r + \mathrm{d}r$ is

$$\frac{\mathrm{d}N}{\mathrm{d}t\,\mathrm{d}r}\,\mathrm{d}r = c\sigma_{\mathrm{T}} n_{\mathrm{ph}} K(r,q)\mathrm{d}r . \tag{12}$$

With the photon energy ϵ_{scat} known, the momentum of the scattered photon has magnitude $\epsilon_{\mathrm{scat}}/c$ and points, for the ultrarelativistic $\gamma \gg 1$ approximation relevant to our simulations, along the direction of the radiating particle's motion. This momentum is subtracted from that of the radiating particle. Furthermore, if ϵ_{scat} is above pair-production threshold with the background photons – that is, if $\epsilon_{\mathrm{scat}}\epsilon_{\mathrm{ph}}/(m_e c^2)^2 \geq 1$ – the scattered photon is promoted to a tracked simulation particle, allowing its later potential absorption to be self-consistently calculated.

Though the Monte Carlo emission scheme described so far relies on the ultrarelativistic approximation $\gamma \gg 1$, some of our simulations contain trans- or even subrelativistic particles. Thus, at lower energies, we need to stitch to a cooling procedure that is non-relativistically valid. To that end, we restrict ourselves to the main regime of astrophysical interest where $\gamma_{\mathrm{KN}} = m_e c^2/4\epsilon_{\mathrm{ph}} \gg 1$ (i.e. $\epsilon_{\mathrm{ph}} \ll 130\,\mathrm{keV}$). This means that there is a broad range of particle energies γ that are ultrarelativistic ($\gamma \gg 1$) but still well below the threshold where quantum Klein–Nishina effects kick in ($\gamma \ll \gamma_{\mathrm{KN}}$): a shared applicability range where either the Monte Carlo emission scheme, which demands $\gamma \gg 1$, or a continuous Thomson radiative drag force, which needs $\gamma \ll \gamma_{\mathrm{KN}}$, could be used. Within this range, we select a threshold 'stitching' particle energy $\gamma_{\text{Th-KN}}$. Above $\gamma_{\text{Th-KN}}$, we employ the IC Monte Carlo emission procedure. Below $\gamma_{\text{Th-KN}}$, we switch to the continuous Thomson radiative drag force (e.g. Blumenthal & Gould 1970; Rybicki & Lightman 1979; Pozdnyakov, Sobol & Syunyaev 1983; for works in the context of relativistic reconnection, cf. Uzdensky 2016; Werner et al. 2019; Mehlhaff et al. 2020, 2021; Sironi & Beloborodov 2020; Sridhar et al. 2021, 2023),

$$\boldsymbol{f}_{\mathrm{rad}} = -(4/3)\sigma_{\mathrm{T}}\gamma^2 U_{\mathrm{ph}}\boldsymbol{\beta}, \tag{13}$$

that enters (via the method of Tamburini et al. 2010) into the particle push through equation (3) and is, importantly, non-relativistically correct. Regarding the choice of $\gamma_{\text{Th-KN}}$, we find that artefacts of the stitching generally disappear when $\gamma_{\text{Th-KN}} \leq 0.1\gamma_{\mathrm{KN}}$ – well inside of the Thomson regime. At the same time, we find that $\gamma_{\text{Th-KN}}$ should be at least of order a few, limiting us to γ_{KN} values that are above 20 or so.

2.2.2 *Monte Carlo pair-production procedure*

We next describe how pair production is processed in the code. For a collision between a gamma-ray and a background photon with angle θ between their velocity vectors and respective energies ϵ_{hard} and ϵ_{ph}, the centre of mass energy is $s = s_0(1-\cos\theta)/2 \leq s_0$, where $s_0 \equiv \epsilon_{\mathrm{hard}}\epsilon_{\mathrm{ph}}/m_e^2 c^4$ must exceed unity for pair production to be possible. Integrating over all possible collision angles θ such that $s > 1$ for a given s_0 yields an overall cross-section presented to a propagating gamma-ray by the background radiation (Gould & Schréder 1967) of[1]

$$\sigma_{\gamma\gamma}(s_0) = \frac{3}{8}\frac{\sigma_{\mathrm{T}}}{s_0^2}\left[\frac{1+\beta_0^2}{1-\beta_0^2}\ln W_0 - \beta_0^2\ln W_0 - \ln^2 W_0 - \frac{4\beta_0}{1-\beta_0^2} + 2\beta_0 + 4\ln W_0\ln(1+W_0) + 4\left(\frac{\pi^2}{12} + \mathrm{Li}_2(-W_0)\right)\right] , \tag{14}$$

where β_0 and W_0 are both s_0-dependent, reading, respectively,

$$\beta_0^2(s_0) \equiv 1 - \frac{1}{s_0} \tag{15}$$

and

$$W_0(s_0) \equiv \frac{1+\beta_0(s_0)}{1-\beta_0(s_0)} . \tag{16}$$

Thus, the pair-production probability accumulated by a gamma-ray of energy ϵ_{hard} in one simulation time-step Δt is

$$p_{\mathrm{abs}} = c n_{\mathrm{ph}} \sigma_{\gamma\gamma}(s_0)\Delta t . \tag{17}$$

[1]The parenthesized term in equation (14), $4\left(\pi^2/12 + \mathrm{Li}_2(-W_0)\right)$, is equal to the last term, $-4L(W_0)$, in equation (10) of Gould & Schréder (1967), but is corrected for the missing factor of 4 in that work.

Because, max $(\sigma_{\gamma\gamma}) \simeq \sigma_{\rm T}/5$, our choice to operate in the regime $p_0 \equiv c\sigma_{\rm T} n_{\rm ph} \Delta t \ll 1$ implies $p_{\rm abs} \leq p_0/5 \ll 1$.

Just as equation (8) determines which potential IC scattering events occur, equation (17) determines which photons produce new pairs. Ordinarily, like the additional steps necessary to determine the final scattered photon energy in the case of IC emission – for example, equation (9) – one would also need to proceed beyond equation (17) to fix the energy of the newborn electron and positron. First, the angle θ, which is integrated out when interested only in the total cross-section (14) and corresponding absorption probability (17), needs to actually be sampled to determine the centre-of-mass energy s. Then, one must also sample the angle of one of the newborn particles' momenta with respect to that of the collision axis in the centre-of-mass frame. We have indeed implemented both steps in ZELTRON, but they are unnecessary when $\epsilon_{\rm hard} \gg m_e c^2 \gg \epsilon_{\rm ph}$ (i.e. when $4\gamma_{\rm KN} \gg 1$), which is all that concerns us in this study. Then, the collision energy budget is entirely dominated by the incoming gamma-ray, and the electron and positron each simply inherit half of the absorbed gamma-ray's momentum.

2.2.3 *Optimization methods for inverse Compton emission*

The Monte Carlo implementation of IC emission sketched in Section 2.2.1 suffers from two performance bottlenecks. First, the loop computing the IC cross-section, equations (7) and (8), for every simulation particle is costly. Second, within this loop, the inversion of the cumulative distribution function in equation (9) is also expensive (though only necessary for the subset of particles that actually scatters photons). We mitigate these issues using two techniques, which we discuss in turn below. We note that, while similar techniques could be applied to the pair-production procedure, our simulations are optically thick to pair production, and the steady-state number of tracked photons is much smaller than the (always growing) number of particles. Thus, the QED physics involving photons, for us, can tolerate a less rigorously optimized implementation.

First, we speed up the assignment of the scattered photon energy – done in the context of Section 2.2.1 through equation (9). To do this, we consider the IC emission in the rest frame of the scattering particle. As we show below, this requires a larger number of random number draws per particle, but enables the use of simpler expressions. These, in turn, furnish an approximation to the scattering cross-section that effectively replaces equation (9), relieving its performance bottleneck. Our handling of the problem this way, presented in detail below, follows closely Levinson & Cerutti (2018) and Crinquand et al. (2020).

Before transforming to the rest frame of the potentially scattering particle, we isolate to an interaction with a single photon rather than with the entire isotropic radiation bath. This is accomplished by first drawing a random photon angle ψ with respect to the direction of the particle's velocity $c\beta$ according to the probability density function

$$p(\psi) \propto 1 - \beta \cos\psi \,, \tag{18}$$

which accounts for the relative lab-frame rate of encounters of the particle with photons incident from different directions. The selected photon is then boosted to the particle's rest frame, where it has energy

$$\epsilon_0' = \gamma \epsilon_{\rm ph} (1 - \beta \cos\psi) \,. \tag{19}$$

Particle rest-frame quantities are primed in our convention.

We now evaluate the spectral (per unit final photon energy) and total (integrated over final photon energies) scattering rate for this interaction in the particle's rest frame. We discuss after this derivation how these two quantities are used by the code. In the primed frame, the scattering cross-section is that of ordinary Compton scattering (e.g. Blumenthal & Gould 1970):

$$\frac{{\rm d}\sigma}{{\rm d}\epsilon_{\rm scat}' {\rm d}\Omega'} = \frac{3\sigma_{\rm T}}{16\pi} \left(\frac{\epsilon_{\rm scat}'}{\epsilon_0'}\right)^2 \left(\frac{\epsilon_0'}{\epsilon_{\rm scat}'} + \frac{\epsilon_{\rm scat}'}{\epsilon_0'} - \sin^2\Psi'\right) \times \delta\left[\epsilon_{\rm scat}' - \frac{\epsilon_0'}{1 + (\epsilon_0'/m_e c^2)(1 - \cos\Psi')}\right] , \tag{20}$$

where Ψ' is the angle through which the photon is scattered in the particle frame. The total scattering rate into final photon energies between $\epsilon_{\rm scat}'$ and $\epsilon_{\rm scat}' + {\rm d}\epsilon_{\rm scat}'$ is then

$$\frac{{\rm d}N}{{\rm d}t {\rm d}\epsilon_{\rm scat}'} = \frac{1}{\gamma} \frac{{\rm d}N}{{\rm d}t' {\rm d}\epsilon_{\rm scat}'} = c n_{\rm ph} \int {\rm d}\Omega' \frac{{\rm d}\sigma}{{\rm d}\epsilon_{\rm scat}' {\rm d}\Omega'} \,, \tag{21}$$

where we used $n_{\rm ph}' = \gamma n_{\rm ph}$. Defining $x \equiv \epsilon_{\rm scat}'/\epsilon_0'$ and $y \equiv m_e c^2/\epsilon_0'$, the delta function only activates for $1/(1 + 2/y) \leq x \leq 1$, in which case the integral evaluates to

$$\frac{{\rm d}N}{{\rm d}t {\rm d}x} = \epsilon_0' \frac{{\rm d}N}{{\rm d}t {\rm d}\epsilon_{\rm scat}'} = \frac{3\sigma_{\rm T}}{8} \times y \left[x + \frac{1}{x} + 2y\left(1 - \frac{1}{x}\right) + y^2\left(\frac{1}{x^2} - \frac{2}{x} + 1\right)\right] . \tag{22}$$

The total scattering rate is then

$$\frac{{\rm d}N}{{\rm d}t} \equiv \int {\rm d}x \frac{{\rm d}N}{{\rm d}t {\rm d}x} = \frac{3\sigma_{\rm T} c n_{\rm ph}}{8} \left[F(1, y) - F\left(\frac{1}{1 + 2/y}, y\right)\right] , \tag{23}$$

where

$$F(x, y) \equiv y \left[-\frac{y^2}{x} + xy(2 + y) + \frac{x^2}{2} + \left(1 - 2y - 2y^2\right) \ln x\right] . \tag{24}$$

These results are used by the code as follows. First, equation (23) is evaluated and multiplied by Δt to determine the probability $p_{\rm emit} = \Delta t \, {\rm d}N/{\rm d}t$ that the scattering occurs during the simulation time-step. This step replaces the evaluation of equation (5). On average, these two procedures are completely equivalent, but in this second method, we have traded the analytic integral over incident photon directions [which yields equation (5)] for a random Monte Carlo sampling over these directions.

Then, for particles that scatter photons, the final photon energy is obtained by drawing a random number $\mathcal{R}$ and inverting the cumulative distribution function: that is, finding x such that

$$\mathcal{R} = \frac{3\sigma_{\rm T} c n_{\rm ph}}{8} \frac{[F(x, y) - F(1/(1 + 2/y), y)]}{{\rm d}N/{\rm d}t} = \frac{F(x, y) - F(1/(1 + 2/y), y)}{F(1, y) - F(1/(1 + 2/y), y)} \,. \tag{25}$$

This replaces the inversion of the equivalent lab-frame cumulative distribution function (9) and, as we now show, constitutes the main advantage of this method. The issue with equation (9) is that its solution cannot be expressed analytically. Instead, the cumulative distribution function (or its inverse) must be stored as a table and consulted for each emitting particle – a costly procedure. However, the function $F(x, y)$ can be approximated, when $x \ll 1$, by only its $\ln x$ term. This approximation breaks down as x approaches unity but, as it turns out, does not very strongly disturb the cumulative distribution function on the right-hand side of equation (25) even when used across all x. If one then approximates

$$F(x, y) \simeq y\left(1 - 2y - 2y^2\right) \ln x \,, \tag{26}$$

the solution to equation (25) can be found as

$$x = \exp\left[-(1-\mathcal{R})\ln(1+2/y)\right]. \tag{27}$$

We have checked that the use of this approximation does not change any discernible aspects of our simulations. We have also conducted experiments (Appendix A) to show that it yields nearly the same average emitted photon energy as the exact cumulative distribution function. It therefore provides a powerful speed-up to the code by facilitating the otherwise impossible analytic evaluation of x without compromising the important physics.

Once the rest-frame scattered photon energy $\epsilon'_{\rm scat} = x\epsilon'_0$ is known, it can be boosted back to the lab frame as follows. First, one notes that, in the ultrarelativistic approximation $\gamma \gg 1$ relevant to this work, the incident photon approaches the particle nearly head-on in the primed frame. Thus, $\cos\Psi' = -\cos\psi'_{\rm scat}$, where $\psi'_{\rm scat}$ is the angle between the emerging photon's momentum and the particle's lab velocity. Then, the delta function in equation (20) can be used to write

$$\cos\psi'_{\rm scat} = -\cos\Psi' = \frac{m_e c^2}{\epsilon'_{\rm scat}} - \frac{m_e c^2}{\epsilon'_0} - 1, \tag{28}$$

which yields the lab-frame emerging photon energy via

$$\epsilon_{\rm scat} = \gamma\epsilon'_{\rm scat}(1+\beta\cos\psi'_{\rm scat}). \tag{29}$$

In addition to considering the IC scattering in the rest frame of each particle, which facilitates the powerful approximations (26) and (27), we also leverage the following second strategy to reduce the cost of the QED module. This standard optimization in the plasma physics literature is known as the *null collision method* (Rees 1968; Lin & Bardsley 1977; Boeuf & Marode 1982; Birdsall 1991). It speeds up the code drastically by avoiding the need to loop over all of the particles during the IC Monte Carlo emission step.

The technique exploits the fact that the probability $p_{\rm emit}$ that an individual particle emits a photon in a given time-step is capped to a global maximum, given by the Thomson limit of equation (5), of $p_0 = c\sigma_{\rm T} n_{\rm ph}\Delta t$. This is a small number in our simulations: of order 10^{-3}. Thus, instead of looping over all of the particles to determine whether fewer than 1 in every $1/p_0$ of them emits a photon, we randomly select a small fraction p_0 of all the particles, loop over this reduced subset, and exactly compensate the limited particle sample by enhancing the per-particle emission probability by the inverse factor $1/p_0$. The null collision method is analogous to the no-time-counter method (Bird 1989; Del Gaudio et al. 2020) except that the latter applies when the collision in question involves two computational particles rather than, as in our case, one computational particle and a fluid field (for us, the prescribed background seed photon bath).

The two techniques described above – the approximation of the cumulative distribution function on IC emission energies in equation (26) and the use of the null collision method – allow us to run QED-PIC simulations with negligible added cost per time-step (of order 10 per cent) taken by the QED module (orange panel in Fig. 1). The main costs are instead the accumulation of particles and photons in the simulation and the typically larger amount of data that one wishes to dump in QED runs. We note, however, that for the regime of reconnection studied in this work, the total number of simulated particles never grows by more than a factor of a few, and thus we do not need particle merging algorithms (Vranic et al. 2015), as implemented in other codes (e.g. OSIRIS, Grismayer et al. 2016; SMILEI, Derouillat et al. 2018; TRISTAN V2, Hakobyan et al. 2019), to regulate such growth at this stage.

3 SIMULATION SETUP

We perform pair-plasma simulations of relativistic collisionless magnetic reconnection using the radiative QED-outfitted (Section 2) electromagnetic PIC code ZELTRON (Cerutti et al. 2013; Cerutti & Werner 2019). The simulations are in a 2D box of physical dimensions $L_x \times L_y = L \times 2L$ and grid size $N_x \times N_y = N \times 2N = 7680 \times 15360$ and are run from $t = 0$ to just over $t = 4L/c$. Spatial dependence is only tracked in the x- and y-directions, but all vectors may have an out-of-plane z-component.

We adopt periodic box boundaries. Each of our simulations can therefore be interpreted as a single reconnection event that completely processes a flux reservoir of macroscopic scale L. This facilitates transparent connections to flaring situations in which reconnection suddenly and explosively liberates the magnetic energy contained in such a flux reservoir, and a major application of our results is to gamma-ray-flaring sources (Section 6). In addition, the periodic boundaries enable a straightforward computation of the pair-plasma yield per reconnection-processed particle (Section 5).

Our four main runs presented in Section 4 share the same setup and parameter values but differ in their modelled radiative effects. In particular, we present one case without any radiative cooling; one with purely continuous Thomson IC cooling (similar to, e.g., Werner et al. 2019; Mehlhaff et al. 2020; Sironi & Beloborodov 2020; Sridhar et al. 2021, 2023); one with fully general IC cooling (including the Klein–Nishina regime) but with pair production artificially turned off; and one with general IC cooling and self-consistent pair production. Synchrotron cooling is ignored in all runs. We describe the non-radiative aspects of our setup in Section 3.1 and move on to the radiative details in Section 3.2.

3.1 Non-radiative aspects of the setup

Our simulations begin with zero electric field. The initial magnetic field is force-free and undergoes reversals via current sheets located at $y_1 = L/2$ and $y_2 = 3L/2$. It has the form (Harrison & Neukirch 2009; Neukirch, Wilson & Harrison 2009)

$$\boldsymbol{B} = B_x(y)\hat{\boldsymbol{x}} + B_z(y)\hat{\boldsymbol{z}} = \pm B_0 \tanh\left(\frac{y-y_{1,2}}{\delta}\right)\hat{\boldsymbol{x}} + B_0\sqrt{\mathrm{sech}^2\left(\frac{y-y_{1,2}}{\delta}\right) + \left(\frac{B_g}{B_0}\right)^2}\,\hat{\boldsymbol{z}}, \tag{30}$$

where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}}$ are unit vectors pointing in the respective x- and z-directions, the $+$ ($-$) sign is taken at y_1 (y_2), and δ is the half-thickness of the current sheets (see also Li et al. 2018; Guo et al. 2021; Goodbred & Liu 2022; Zhang et al. 2022; French et al. 2023, for a few relativistic reconnection studies employing the same initial magnetic field). In the reconnection upstream region far away from the current sheets ($|y - y_{1,2}| \gg \delta$), equation (30) reduces to a uniform field with in-plane component $\pm B_0\hat{\boldsymbol{x}}$ and out-of-plane guide-field component $B_g\hat{\boldsymbol{z}} = 0.15B_0\hat{\boldsymbol{z}}$. In addition to this modest guide field, which accompanies the upstream plasma into the reconnection layer, there is also a strong localized ($|y - y_{1,2}| \ll \delta$) initial out-of-plane field of peak strength $\sqrt{B_0^2 + B_g^2}$ that supplies the force-free magnetic field reversal. Both this strong localized out-of-plane field and the upstream guide field provide some pressure support to the plasma energized by reconnection as it radiatively cools down, which helps ensure the Debye length in the simulations remains well resolved.

Our initial fields satisfy the force-free condition, $\boldsymbol{J}\times\boldsymbol{B}/c = -\boldsymbol{\nabla}\boldsymbol{B}^2/8\pi + \boldsymbol{B}\cdot\boldsymbol{\nabla}\boldsymbol{B}/4\pi = 0$. Thus, and unlike the case of a Harris

equilibrium (Kirk & Skjæraasen 2003), no added plasma pressure is needed inside the initial current layers. This allows us to start the simulations with a plasma of completely homogeneous initial (electron + positron) number density, n_0, and temperature, $T_0 = \theta_0 m_e c^2 = 24 m_e c^2$. Specifically, we use a relativistic Maxwell–Jüttner initial plasma distribution function. This setup avoids spurious pair production from radiation emitted by the hot particles that would be needed to supply the initial pressure support in a Harris current sheet.[2]

The number density, n_0, and initial reconnecting field strength, B_0, together define the *cold magnetization*,

$$\sigma_{\rm c,0} \equiv \frac{B_0^2}{4\pi n_0 m_e c^2}\,, \tag{31}$$

equal to about twice the reconnecting magnetic field energy per particle. While $\sigma_{\rm c,0}$ defines how much energy a typical reconnection-accelerated particle may acquire, another similar quantity, the *hot magnetization*,

$$\sigma_{\rm h,0} \equiv \frac{B_0^2}{4\pi w_0}\,, \tag{32}$$

defines the magnetic dominance of the upstream region. Here, the initial plasma enthalpy density, w_0, can be written as $w_0 = u_0 + P_0$, where u_0 and P_0 are, respectively, the initial plasma internal energy density and pressure. For a non-relativistically cold initial temperature, $\theta_0 \ll 1$, the enthalpy density is dominated by rest-mass energy, $w_0 \simeq u_0 = n_0 m_e c^2$, and, hence, $\sigma_{\rm h,0} \simeq \sigma_{\rm c,0}$. For a relativistically hot plasma, $\theta_0 \gg 1$, the thermal kinetic energy and pressure dominate the enthalpy, $w_0 \simeq 4P_0 = 4\theta_0 n_0 m_e c^2$, and, in this case, $\sigma_{\rm h,0} \simeq \sigma_{\rm c,0}/4\theta_0 = 1/(2\beta_{\rm pl})$, where $\beta_{\rm pl}$ is plasma beta. For all regimes, $\sigma_{\rm h,0} \leq \sigma_{\rm c,0}$, meaning that $\sigma_{\rm h,0} \gg 1$ is a stronger condition than $\sigma_{\rm c,0} \gg 1$. In fact, because $\sigma_{\rm h,0}$ sets the plasma Alfvén speed $v_A = c\sqrt{\sigma_{\rm h,0}/(1+\sigma_{\rm h,0})}$, a high $\sigma_{\rm h,0}$ means that the energy budget permits not just relativistic individual particles (which merely requires high $\sigma_{\rm c,0}$), but also for the collective bulk motion of the plasma itself to become highly relativistic. In our simulations, we set $\sigma_{\rm c,0} = 1.2 \times 10^5$ and $\sigma_{\rm h,0} = \sigma_{\rm c,0}/4\theta_0 = 1250$.

Though the force-free initial condition does not require any extra plasma density or temperature in the current sheets to balance the upstream magnetic pressure, the field-reversing currents must still be supplied. Correspondingly, we set a local fraction

$$f_{\rm d}(y) = \mathrm{sech}(\tilde{y}/\delta)\sqrt{\frac{1+(B_g/B_0)^2}{1+[B_g\cosh(\tilde{y}/\delta)/B_0]^2}} \tag{33}$$

of the positrons in motion at a drift velocity

$$c\boldsymbol{\beta}_{\rm d,i}(y) = \frac{c\beta_{\rm d}}{\sqrt{1+(B_g/B_0)^2}} \times \left[\mp\sqrt{\mathrm{sech}^2(\tilde{y}/\delta)+(B_g/B_0)^2}\,\hat{\boldsymbol{z}} - \tanh(\tilde{y}/\delta)\hat{\boldsymbol{x}}\right] \tag{34}$$

so that they carry half the field-reversing current

$$\boldsymbol{J} = c\boldsymbol{\nabla}\times\boldsymbol{B}/4\pi = ecn_0 f_{\rm d}(y)\boldsymbol{\beta}_{\rm d,i}(y)\,, \tag{35}$$

where the $\mp$ corresponds to equation (30), e is the positron charge, and $\tilde{y} \equiv y - y_{1,2}$. The other half of the current is carried by the initial electrons, which are assigned $\boldsymbol{\beta}_{\rm d,e} = -\boldsymbol{\beta}_{\rm d,i}$. The drifting particles follow a drifting Maxwell–Jüttner distribution function with initial temperature, T_0, defined in their local [boosted by $\boldsymbol{\beta}_{\rm d,i/e}(y)$] rest frame.

The force-free setup ties the current sheet half-thickness δ to the other length-scales in the problem as follows. The drifting plasma supplies a current density $ecn_0\beta_{\rm d} \leq ecn_0$, whereas the current needed at the heart of the layer is, according to Ampère's law (35), $cB_0/4\pi\delta$. This means that

$$\delta = \frac{B_0}{4\pi e n_0 \beta_{\rm d}} = \frac{\sigma_{\rm c,0}\,\rho_0}{\beta_{\rm d}} \geq \sigma_{\rm c,0}\,\rho_0\,, \tag{36}$$

where we have introduced the *nominal gyroradius*,

$$\rho_0 \equiv \frac{m_e c^2}{eB_0}\,. \tag{37}$$

Equation (36) demands that the current sheet half-thickness be limited to the typical gyroradius, $\sigma_{\rm c,0}\,\rho_0$, of reconnection-energized particles. Therefore, in order to achieve a high aspect ratio, L/δ, while also complying with the other demanding radiative constraints described later (Section 3.2), we assign $\beta_{\rm d} = 0.9$. This corresponds to $\delta = 1.1\sigma_{\rm c,0}\rho_0$ and to an initial drifting-particle Lorentz factor of 2.3 (still much less than θ_0).

The current sheet aspect ratio, $L/\delta = L\beta_{\rm d}/\sigma_{\rm c,0}\,\rho_0 \simeq L/\sigma_{\rm c,0}\,\rho_0$, represents not just a ratio of length-scales but also one of particle energy scales. This is because, during reconnection, some particles are swept into the vicinity of an X-point (X-line in 3D), which is a region wherein the in-plane magnetic field reconnects. There, they become unmagnetized and linearly accelerated by the out-of-plane reconnection electric field, $E_{\rm rec} = \beta_{\rm rec}B_0 \simeq 0.1B_0$, where $\beta_{\rm rec} \equiv 0.1v_A/c \simeq 0.1$ is the fiducial collisionless reconnection rate. As described by Werner et al. (2016), in systems with modest aspect ratios, for which the reconnection layer is dominated by a single X-point, particles can be accelerated by $E_{\rm rec}$ until, after travelling an out-of-plane distance $\sim L$, they escape the acceleration region with final *system-size-limited Lorentz factor*

$$\gamma_{\rm max} \equiv \frac{0.1eB_0L}{m_e c^2} = \frac{0.1L}{\rho_0}\,. \tag{38}$$

However, for larger systems, the reconnection layer tears into a hierarchical chain of plasma-filled magnetic islands, or *plasmoids*, studded with many X-points. Then, particles cannot travel the whole system size before escaping a given acceleration zone. Instead, at least in 2D, they may travel a distance comparable to that between the smallest-scale plasmoids, which subsequently capture the particles, limiting them to energies of order the intrinsic reconnection *X-point acceleration Lorentz factor* (cf. Sironi et al. 2016; Werner et al. 2016; Uzdensky 2022),

$$\gamma_{\rm X} \equiv 4\sigma_{\rm c,0}\,. \tag{39}$$

The aspect ratio becomes large enough that the X-point acceleration channel is intrinsically limited by the self-consistent evolution of the plasmoid chain rather than by the modest size of the system when these two energy scales cross each other: $\gamma_{\rm max}/\gamma_{\rm X} \geq 1 \Rightarrow L/\sigma_{\rm c,0}\,\rho_0 \geq 40$ (Werner et al. 2016). Thus, the requirement to have a large aspect ratio $L/\delta \gg 1$ is synonymous with having a healthy separation between $\gamma_{\rm max}$ and the particle energies, $\sim\gamma_{\rm X}$, at which reconnection X-point acceleration saturates. We adopt $\gamma_{\rm max} = 9.1\sigma_{\rm c,0}$, which corresponds to $L = 91\sigma_{\rm c,0}\,\rho_0$. This meets the fiducial $\gamma_{\rm max} > \gamma_{\rm X}$ criterion but, for reasons described below, yields $\gamma_{\rm max}/\sigma_{\rm c,0}$ smaller than typical for simulations of the numerical size, $N_x \times N_y = 7680 \times 15360$, that we present.

[2]While such pair production could also be avoided in a traditional Harris equilibrium by specifying a high plasma overdensity in the initial current layer, this would reduce the initial Debye length of the layer plasma, and would thus create problems for the Debye length resolution in our simulations.

Table 1. Simulation parameters used in this study. Non-radiative parameters (described in Section 3.1) are presented in the upper part of the table and radiative parameters (described in Section 3.2) below. The radiative parameters $\gamma_{\rm cool}$ and $\gamma_{\rm rad,\,T}$ apply only to the three simulations with radiative cooling, $\gamma_{\rm KN}$ applies only to the two simulations with fully general IC cooling (including Klein–Nishina effects), and $\tau_{\gamma\gamma}$ applies only to the one simulation with pair production. The expression for $\sigma_{\rm h,0}$ assumes $\theta_0 \gg 1$.

Parameter	Symbol (= definition)	Value (= equivalent)
Upstream reconnecting field strength	B_0	
Nominal gyroradius	$\rho_0 = m_e c^2/eB_0$	
Initial upstream density	n_0	
Initial cold magnetization	$\sigma_{\rm c,0} = B_0^2/4\pi n_0 m_e c^2$	120 000
Initial hot magnetization	$\sigma_{\rm h,0} = B_0^2/16\pi n_0 \theta_0 m_e c^2$	1250
Initial upstream temperature	$\theta_0 = T_0/m_e c^2$	$2\times10^{-4}\sigma_{\rm c,0} = 24$
System size	$L_x = L$	$91\sigma_{\rm c,0}\rho_0$
System-size-limited Lorentz factor	$\gamma_{\rm max} = 0.1L/\rho_0$	$9.1\sigma_{\rm c,0} = 1.1\times10^6$
Guide field	B_g	$0.15B_0$
Layer drift velocity	$\beta_{\rm d}c$	$0.9c$
Layer half-thickness	$\delta = \sigma_{\rm c,0}\rho_0/\beta_{\rm d}$	$1.1\sigma_{\rm c,0}\rho_0$
Cell size	$\Delta x,\ \Delta y$	$\sigma_{\rm c,0}\rho_0/85$
Time-step	Δt	$3\Delta x/4\sqrt{2}c$
Grid size	$N_x = N$	7680
Computational particles per cell		20
Soft photon energy density	$U_{\rm ph}$	
Soft photon energy	$\epsilon_{\rm ph}$	
Soft photon number density	$n_{\rm ph} = U_{\rm ph}/\epsilon_{\rm ph}$	
Nominal cooling Lorentz factor	$\gamma_{\rm cool} = 3m_e c^2/4U_{\rm ph}\sigma_{\rm T}L$	$3.6\times10^{-3}\sigma_{\rm c,0} = 430$
Thomson IC-limited Lorentz factor	$\gamma_{\rm rad,\,T} = (0.3eB_0/4\sigma_{\rm T}U_{\rm ph})^{1/2} = (\gamma_{\rm max}\gamma_{\rm cool})^{1/2}$	$0.18\sigma_{\rm c,0} = 2.2\times10^4$
Klein–Nishina Lorentz factor	$\gamma_{\rm KN} = m_e c^2/4\epsilon_{\rm ph}$	$0.046\sigma_{\rm c,0} = 5500$
Box pair-production optical depth	$\tau_{\gamma\gamma} = n_{\rm ph}\sigma_{\rm T}L/5 = 3\gamma_{\rm KN}/5\gamma_{\rm cool}$	7.7

Our grid resolution $\Delta x = \Delta y$ is set by the need to resolve the initial Debye length,

$$\lambda_{\rm D,0} = \sqrt{\frac{\theta_0 m_e c^2}{4\pi e^2 n_0}} = \sqrt{\theta_0 \sigma_{\rm c,0}}\,\rho_0\,, \tag{40}$$

and we set $\Delta x = \lambda_{\rm D,\,0}/1.2$ in all runs. Because we operate in the highly relativistic regime, $\sigma_{\rm c,0} \geq \sigma_{\rm h,0} \gg 1$, the Debye length is much smaller than the typical energized particle's gyroradius, $\sigma_{\rm c,0}\rho_0 \sim \delta$. This is largely why we cannot afford a larger $\gamma_{\rm max}/\sigma_{\rm c,0}$ – our choice of $\sigma_{\rm h,0} = 1250$ induces a large separation between the plasma microscales, $\sigma_{\rm c,0}\,\rho_0/\lambda_{\rm D,0} = \sqrt{\sigma_{\rm c,0}/\theta_0} \simeq 2\sqrt{\sigma_{\rm h,0}} \sim 70$, which occupies much of our grid resolution to resolve.[3] We do, however, under-resolve the gyroradii of particles with energies less than $\gamma_{\Delta x} \equiv (\theta_0\sigma_{\rm c,0})^{1/2}/1.2$, including the upstream particles (energies $\sim\theta_0 \ll \gamma_{\Delta x}$). We do not observe any strong artefacts of this in our results, and the energy in our simulations is conserved to 1 per cent or better except where noted (i.e. in Table 2).

Given the cell size, Δx, we employ a time-step $\Delta t = 3\Delta x/4\sqrt{2}c \simeq 0.5\Delta x/c$: slightly smaller than the maximum allowed Courant–Friedrichs–Lewy step $\Delta x/\sqrt{2}c$. While this does not introduce significant numerical dispersion (see fig. 2 of Greenwood et al. 2004), it typically improves the simulations' energy conservation by about 0.5 to 1 percentage points (see, e.g., the simulations of Table 2 that employ a larger time-step and exhibit slightly poorer energy conservation). We initially place 20 computational particles (electrons + positrons) per cell, except within a few δ of $y_{1,2}$ where we place 40 per cell. The particle currents are deposited to the grid using the charge-conserving method of Esirkepov (2001).

[3]While $\sigma_{\rm c,0}\,\rho_0/\lambda_{\rm D,0} = \sqrt{\sigma_{\rm c,0}/\theta_0}$ is generally true, the approximation, $\sqrt{\sigma_{\rm c,0}/\theta_0} \simeq 2\sqrt{\sigma_{\rm h,0}}$, assumes $\theta_0 \gg 1$.

The principal non-radiative simulation parameters described in this section and their values are summarized in the top part of Table 1.

3.2 Radiative aspects of the setup

The two principle radiative parameters in our simulations are the total energy density, $U_{\rm ph}$, and monochromatic single-photon energy, $\epsilon_{\rm ph}$, of the IC seed photons [equation (4)]. These photons are not tracked, but provide a static, homogeneous, isotropic target population for the charged particles to scatter. Though only the two numbers $U_{\rm ph}$ and $\epsilon_{\rm ph}$ need to be prescribed to fully specify the radiative physics, they introduce a variety of derived energy- and length-scales into the problem, and, hence, divide the parameter space into many different regimes. We first provide a brief overview of these regimes in Section 3.2.1, summarizing the more detailed discussion from our earlier work, Mehlhaff et al. (2021). Afterwards, in Section 3.2.2, we discuss how these parameters are chosen for our simulations (as in Table 1).

3.2.1 Reconnection scales introduced by radiative physics

For particles with Lorentz factors $\gamma \ll \gamma_{\rm KN} \equiv m_e c^2/4\epsilon_{\rm ph}$, IC emission proceeds in the Thomson regime, where typical scattered photons are enhanced to energies $\sim\gamma^2\epsilon_{\rm ph}$, up to a maximum of $4\gamma^2\epsilon_{\rm ph}$. Because no individual photon robs the particle of a significant fraction of its energy, cooling proceeds continuously and is modelled (Section 2.1) as a radiative drag force, $\boldsymbol{f}_{\rm rad}$. The total power radiated by a given particle in the Thomson regime is [cf. equation (13)]

$$P_{\rm T}(\gamma) = |c\boldsymbol{f}_{\rm rad}\cdot\boldsymbol{\beta}| = \frac{4}{3}\sigma_{\rm T}c\gamma^2\beta^2 U_{\rm ph}\,, \tag{41}$$

where $c\boldsymbol{\beta}$ is the 3-velocity of the particle.

However, at sufficiently high γ, the maximum Thomson emission energy, $4\gamma^2\epsilon_{\rm ph}$, rivals the scattering particle's energy $\gamma m_e c^2$ (the two

are formally equal at $\gamma_{\rm KN} = m_e c^2/4\epsilon_{\rm ph}$). Then, particles pass into the Klein–Nishina regime where they lose energy in discrete quanta, $\sim\gamma m_e c^2$. Here, the cross-section [e.g. equations (7) and (20)] needs to be described by QED and gives rise to an average power radiated per particle of

$$P_{\rm IC}(\gamma) = P_{\rm T}(\gamma) f_{\rm KN}(\gamma/\gamma_{\rm KN}) , \tag{42}$$

where $f_{\rm KN}(q)$, with $q = \gamma/\gamma_{\rm KN}$, is the dimensionless function (cf. Jones 1968; Nalewajko, Yuan & Chruślińska 2018; Mehlhaff et al. 2021)

$$f_{\rm KN}(q) = \frac{9}{q^3}\left[\left(\frac{q}{2}+6+\frac{6}{q}\right)\ln(1+q) - \frac{1}{(1+q)^2}\left(\frac{11}{12}q^3+6q^2+9q+4\right) - 2 + 2\,{\rm Li}_2(-q)\right] \tag{43}$$

and ${\rm Li}_2(q)$ is the dilogarithm. As necessary to recover the Thomson limit, $f_{\rm KN}(q \ll 1) \to 1$. In the opposite, deep Klein–Nishina regime, $f_{\rm KN}(q \gg 1) \simeq (9/2q^2)[\ln(q) - 11/6]$. Equations (41) and (42) define the respective Thomson-limit and general-case IC cooling times,

$$t_{\rm cool,T}(\gamma) = \frac{\gamma m_e c^2}{P_{\rm T}(\gamma)} \simeq \frac{3 m_e c}{4\sigma_{\rm T} U_{\rm ph}\gamma} = \frac{L}{c}\frac{\gamma_{\rm cool}}{\gamma} \tag{44}$$

and

$$t_{\rm cool,IC}(\gamma) = \frac{\gamma m_e c^2}{P_{\rm IC}(\gamma)} \simeq \frac{3 m_e c}{4\sigma_{\rm T} U_{\rm ph}\gamma\, f_{\rm KN}(\gamma/\gamma_{\rm KN})} = \frac{L}{c}\frac{\gamma_{\rm cool}}{\gamma\, f_{\rm KN}(\gamma/\gamma_{\rm KN})} , \tag{45}$$

where we make the relativistic $\beta \simeq 1$ approximation in both cases. The above expressions also invoke the *nominal efficient-cooling Lorentz factor*

$$\gamma_{\rm cool} \equiv \frac{3 m_e c^2}{4\sigma_{\rm T} U_{\rm ph} L} , \tag{46}$$

defined by $t_{\rm cool,T}(\gamma_{\rm cool}) \equiv L/c$. Thus, in the Thomson regime, $\gamma_{\rm cool}$ corresponds to the minimum Lorentz factor for a particle to cool on time-scales shorter than the system light/Alfvén-crossing time, L/c. One may express $\gamma_{\rm cool}$ in terms of the radiative compactness, $\ell \equiv U_{\rm ph}\sigma_{\rm T} L/m_e c^2$, as $\gamma_{\rm cool} = 3/4\ell$. Unlike a real particle Lorentz factor, the formal parameter $\gamma_{\rm cool}$ can be less than 1. This corresponds to the highly compact case, $\ell > 1$, and signals that all particles cool to non-relativistic energies in less than L/c.

In addition to $\gamma_{\rm cool}$ – and ignoring Klein–Nishina effects for the moment – another radiative Lorentz factor scale may be defined by equating the radiative cooling time, $t_{\rm cool,T}(\gamma)$, to the linear acceleration time for particles experiencing the reconnection electric field $E_{\rm rec} = 0.1 B_0$ near reconnection X-points,

$$t_{\rm X}(\gamma) = \frac{\gamma m_e c^2}{0.1 e c B_0} = \frac{10\gamma\rho_0}{c} . \tag{47}$$

Putting $t_{\rm X}(\gamma_{\rm rad,T}) \equiv t_{\rm cool,T}(\gamma_{\rm rad,T})$ yields the *nominal radiatively-limited Lorentz factor* (cf. Nalewajko 2016; Uzdensky 2016; Werner et al. 2019; Mehlhaff et al. 2020, 2021; Sironi & Beloborodov 2020; Sridhar et al. 2021, 2023),

$$\gamma_{\rm rad,T} \equiv \sqrt{\frac{0.3 e B_0}{4\sigma_{\rm T} U_{\rm ph}}} . \tag{48}$$

A particle with Lorentz factor $\gamma > \gamma_{\rm rad,T}$ experiences a Thomson radiative drag force stronger than the acceleration force from the reconnection electric field. Exceeding this energy is impossible in the Thomson regime in the absence of other, faster (i.e. operating on time-scales $< t_{\rm X}$) acceleration mechanisms (Mehlhaff et al. 2021).

The Lorentz factors $\gamma_{\rm cool}$ and $\gamma_{\rm rad,T}$ are not independent; they are both set by the same underlying radiative parameter $U_{\rm ph}$. Hence, $\gamma_{\rm rad,T}$ is fixed by $\gamma_{\rm cool}$ and $\gamma_{\rm max}$ (Mehlhaff et al. 2021):

$$\gamma_{\rm rad,T}^2 = \gamma_{\rm cool}\gamma_{\rm max} . \tag{49}$$

Equation (49) reflects the fact that, since the X-point acceleration time – which defines $\gamma_{\rm rad,T}$ – is faster than the system light crossing time – which defines $\gamma_{\rm cool}$ – one generally has $\gamma_{\rm cool} < \gamma_{\rm rad,T} < \gamma_{\rm max}$. The exception is the non-radiative regime, where $\gamma_{\rm cool} > \gamma_{\rm rad,T} > \gamma_{\rm max}$ and all particles cool on time-scales exceeding L/c.

The Lorentz factors $\gamma_{\rm cool}$, $\gamma_{\rm rad,T}$, and $\gamma_{\rm max}$ characterize the topology of the time-scales L/c, $t_{\rm X}(\gamma)$, and $t_{\rm cool,T}(\gamma)$ by defining all three pairwise intersection points in γ-space: $t_{\rm cool,T}(\gamma_{\rm cool}) \equiv L/c$, $t_{\rm X}(\gamma_{\rm rad,T}) \equiv t_{\rm cool,T}(\gamma_{\rm rad,T})$, and [using equations (38) and (47)] $t_{\rm X}(\gamma_{\rm max}) \equiv L/c$. We depict this graphically in Fig. 2. Such a topological view helps illustrate a noteworthy point of conceptual consistency. Namely, because L/c, $t_{\rm X}(\gamma)$, and $t_{\rm cool,T}(\gamma)$ depend, respectively, on L, B_0, and $U_{\rm ph}$, the following hold: $\gamma_{\rm cool}$, at the intersection of $t_{\rm cool,T}(\gamma)$ and L/c, depends on $U_{\rm ph}$ and L but not on B_0; $\gamma_{\rm rad,T}$, at the intersection of $t_{\rm cool,T}(\gamma)$ and $t_{\rm X}(\gamma)$, depends on $U_{\rm ph}$ and B_0 but not on L; and $\gamma_{\rm max}$, at the intersection of $t_{\rm X}(\gamma)$ and L/c, depends on B_0 and L but not on $U_{\rm ph}$ (it is the only non-radiative parameter of the three).

Let us now summarize the situation for Thomson IC cooling. In this regime, according to equation (41), cooling depends only on $U_{\rm ph}$. The seed photon energy, $\epsilon_{\rm ph}$, completely drops out of the dynamics (though it still influences the observed photon energies). The parameter $U_{\rm ph}$ can then be recast in terms of the particle energy scales $\gamma_{\rm cool}$ and $\gamma_{\rm rad,T}$, which characterize intersection points of important time-scales in the reconnection problem (Fig. 2).

To illuminate the influence of the seed photon energy, $\epsilon_{\rm ph}$, in the general IC case (i.e. including the Klein–Nishina regime), a similar procedure can be employed as for the Thomson limit. Here, the relevant energy scale in terms of which $\epsilon_{\rm ph}$ is recast, $\gamma_{\rm KN}$, has already been introduced; it is the Lorentz factor above which particles lose their energies in discrete photon quanta. We then only need to repeat the comparison of time-scales as done above, but now replacing the Thomson-regime $t_{\rm cool,T}(\gamma)$ with the more general $t_{\rm cool,IC}(\gamma)$. This is done graphically in Fig. 2, where, in addition to the time-scales L/c, $t_{\rm X}(\gamma)$, and $t_{\rm cool,T}(\gamma)$ discussed above, we also plot $t_{\rm cool,IC}(\gamma)$ for three representative values of $\gamma_{\rm KN}$: $\gamma_{\rm KN}^{\rm (lo)}$, $\gamma_{\rm KN}^{\rm (med)}$, and $\gamma_{\rm KN}^{\rm (hi)}$. In the figure, we define the new auxiliary energy scale $\gamma_{\rm rad,IC}$ as the generalization of $\gamma_{\rm rad,T}$ to include Klein–Nishina effects: $t_{\rm cool,IC}(\gamma_{\rm rad,IC}) \equiv t_{\rm X}(\gamma_{\rm rad,IC})$. As is the case for $\gamma_{\rm rad,T}$, exceeding $\gamma_{\rm rad,IC}$ is impossible barring acceleration channels faster than direct acceleration near reconnection X-points (Mehlhaff et al. 2021).

The three sets of curves in Fig. 2 corresponding to the different $\gamma_{\rm KN}$ values illustrate the three time-scale topologies that may be realized as $\gamma_{\rm KN}$ is brought in from infinity (equivalently, as the seed photon energy, $\epsilon_{\rm ph}$, is increased from zero). First, when $\gamma_{\rm KN} > \gamma_{\rm rad,T}$ (Fig. 2, green $\gamma_{\rm KN}^{\rm (hi)}$ curves) the generalized radiative Lorentz factor, $\gamma_{\rm rad,IC}$, remains approximately equal to $\gamma_{\rm rad,T}$. Thus, radiative losses inhibit particles from accessing energies $\gtrsim\gamma_{\rm KN}$ where they would experience Klein–Nishina effects. This regime therefore essentially reduces to the Thomson limit of radiative reconnection.

The next regime occurs once $\gamma_{\rm KN}$ falls below $\gamma_{\rm rad,T}$ (Fig. 2, purple $\gamma_{\rm KN}^{\rm (med)}$ curves). Then, Klein–Nishina effects [entering through

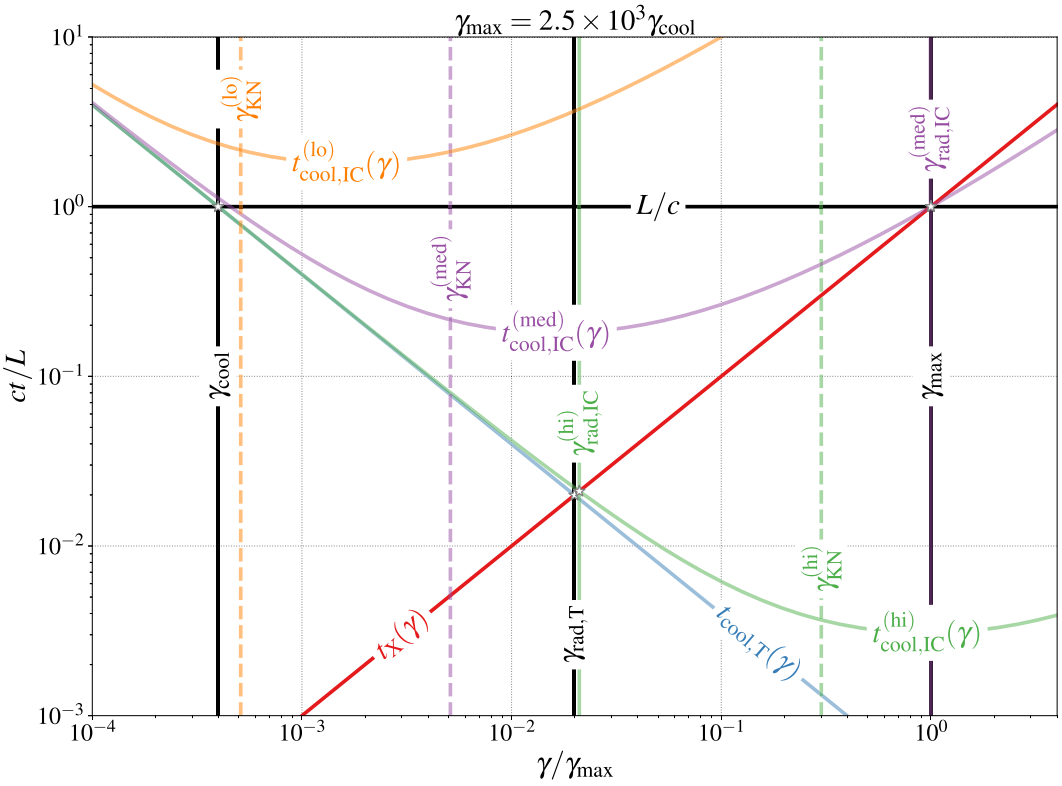

Figure 2. Important reconnection time-scales plotted as functions of γ. Intersection points define characteristic Lorentz factors. The isosceles triangle formed by the non-radiative time-scales, L/c (horizontal solid black line) and $t_{\rm X}(\gamma)$ [equation (47); red line], with the Thomson IC cooling time, $t_{\rm cool,\,T}(\gamma)$ [equation (44), blue line], yields the condition $\gamma_{\rm rad,T}^2 = \gamma_{\rm cool}\gamma_{\rm max}$ [equation (49)]. Klein–Nishina effects enter through finite $\gamma_{\rm KN}$ [equation (7)] and are captured by replacing $t_{\rm cool,\,T}(\gamma)$ with the general IC cooling time, $t_{\rm cool,\,IC}(\gamma)$ [equation (45), where $t_{\rm cool,T}(\gamma) = \lim_{\gamma_{\rm KN}\to\infty} t_{\rm cool,IC}(\gamma)$]. We depict three illustrative values of $\gamma_{\rm KN}$, $\gamma_{\rm KN}^{\rm (lo)} < \gamma_{\rm KN}^{\rm (med)} < \gamma_{\rm KN}^{\rm (hi)}$, corresponding, respectively, to the orange, purple, and green sets of curves. These exemplify the three main topologies, described in Section 3.2.1, that can be formed by $t_{\rm cool,\,IC}(\gamma)$ with the non-radiative time-scales, L/c and $t_{\rm X}(\gamma)$. The depicted $\gamma_{\rm cool}$, $\gamma_{\rm rad,\,T}$, and $\gamma_{\rm max}$ equal those in the runs of Section 4 with radiative cooling: IC(Th), IC(KN), and IC(KN)+PP. In addition, while $\gamma_{\rm KN}^{\rm (lo)}$ and $\gamma_{\rm KN}^{\rm (hi)}$ are chosen arbitrarily for illustrative purposes, $\gamma_{\rm KN}^{\rm (med)} \simeq 13\gamma_{\rm cool}$ matches the runs of Section 4 with Klein–Nishina effects: IC(KN) and IC(KN)+PP.

the function $f_{\rm KN}(\gamma/\gamma_{\rm KN})$ falling appreciably below unity] create a pronounced departure of $t_{\rm cool,\,IC}(\gamma)$ from the Thomson limit. As a result, $\gamma_{\rm rad,\,IC}$ rises rapidly (it depends super-exponentially on $\gamma_{\rm rad,\,T}/\gamma_{\rm KN} > 1$; Mehlhaff et al. 2021) until it eventually crosses $\gamma_{\rm max}$. The $\gamma_{\rm KN}^{\rm (med)}$ curves of Fig. 2 illustrate the particular case where $\gamma_{\rm rad,\,IC} = \gamma_{\rm max}$, which already occurs when $\gamma_{\rm KN}$ is smaller than $\gamma_{\rm rad,\,T}$ by just a factor of a few. As $\gamma_{\rm KN}$ continues to diminish from this point, $\gamma_{\rm rad,\,IC}$ comes to exceed $\gamma_{\rm max}$, meaning that Klein–Nishina effects effectively remove the radiative limit on direct X-point acceleration (though of course the intrinsic and system-size limits, $\gamma_{\rm X}$ and $\gamma_{\rm max}$, are still present). Nevertheless, as long as $\gamma_{\rm KN} \gtrsim \gamma_{\rm cool}$, there remains a range of particle energies characterized by relatively fast cooling, with $t_{\rm cool,\,IC}(\gamma) < L/c$.

Finally, as $\gamma_{\rm KN}$ approaches $\gamma_{\rm cool}$ (Fig. 2, orange $\gamma_{\rm KN}^{\rm (lo)}$ curves), the $t_{\rm cool,\,IC}(\gamma)$ curve is lifted above the line L/c for all γ. Here, Klein–Nishina effects suppress cooling to the point that all particles cool on times longer than L/c, an effectively non-radiative regime.

Of these three cases, the one where Klein–Nishina effects are expected to influence the reconnection dynamics is the second; only there can particles access energies $>\gamma_{\rm KN}$ while maintaining a relatively rapid cooling time. We hence call this the *Klein–Nishina radiative regime*, characterized by the scale hierarchy

$$\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,T} < \gamma_{\rm max}\,. \tag{50}$$

As shown next, this hierarchy is not just important from the standpoint of radiative cooling; it is also where pair production from the emitted photons may play an important role.

Particles with Lorentz factor $\gamma_{\rm KN}$ scatter photons to typical energies $\epsilon_{\rm scat} \sim \gamma_{\rm KN} m_e c^2 = (m_e c^2)^2/4\epsilon_{\rm ph}$: that is, close to pair-production threshold with the seed-photon background. This means that particles must necessarily be accelerated above $\gamma_{\rm KN}$ in order for a significant fraction of their radiated energy to be recaptured as fresh pairs. In addition, the fiducial pair-production optical depth is [using the peak cross-section $\sigma_{\gamma\gamma} \simeq \sigma_{\rm T}/5$; equation (14); see also Mehlhaff et al. (2021)]

$$\tau_{\gamma\gamma} = \frac{U_{\rm ph}\sigma_{\rm T} L}{5\epsilon_{\rm ph}} = \frac{3\gamma_{\rm KN}}{5\gamma_{\rm cool}}\,, \tag{51}$$

meaning that placing $\gamma_{\rm KN}$ between $\gamma_{\rm cool}$ and $\gamma_{\rm rad,\,T}$ enables both the emission of above-threshold gamma-rays and their absorption inside the system on time-scales $L/c\tau_{\gamma\gamma} < L/c$.

In summary, the Klein–Nishina radiative scale hierarchy, $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, triggers three simultaneous and important QED effects:

(i) it permits particles to reach energies $>\gamma_{\rm KN}$, where their IC cooling transitions to the quantized Klein–Nishina regime;

(ii) it permits particles to reach energies $>\gamma_{\rm KN}$, where their IC-emitted gamma-rays exceed pair-production threshold with the background radiation bath; and

(iii) it guarantees almost all of these gamma-rays to be absorbed inside the system on subdynamical ($<L/c$) time-scales, allowing the resulting pairs to feed back on reconnection.

This is the target regime of this study, and how we realize it in numerical simulations is the topic of the next section.

3.2.2 *Selection of the radiative reconnection parameters*

We now discuss our choices of the radiative reconnection parameters, $U_{\rm ph}$ and $\epsilon_{\rm ph}$, recast, as described in the preceding Section 3.2.1, in terms of the energy scales $\gamma_{\rm cool}$, $\gamma_{\rm rad,\,T}$, and $\gamma_{\rm KN}$, plus the pair-production optical depth $\tau_{\gamma\gamma}$. We set as a first goal the Klein–Nishina scale hierarchy, $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, discussing afterward the necessary placement of the remaining non-radiative energy scale, $\sigma_{\rm c,0}$ (or, equivalently, $\gamma_{\rm X} \equiv 4\sigma_{\rm c,0}$) within this base ordering. A key issue is that as much space as possible needs to be opened up between each successive energy scale in the Klein–Nishina hierarchy. The reason for this is that a high optical depth, $\tau_{\gamma\gamma} = 3\gamma_{\rm KN}/5\gamma_{\rm cool}$, demands that $\gamma_{\rm KN}$ substantially exceed $\gamma_{\rm cool}$. At the same time, $\gamma_{\rm KN}$ cannot be as large as $\gamma_{\rm rad,\,T}$ without, as previously described, leading to a prohibitively small radiatively limited Lorentz factor, $\gamma_{\rm rad,\,IC}$. Thus, the gaps from $\gamma_{\rm cool}$ to $\gamma_{\rm KN}$ and, then, from $\gamma_{\rm KN}$ to $\gamma_{\rm rad,\,T}$ both need to be as wide as resolution requirements permit. In effect, the separation from $\gamma_{\rm cool}$ to $\gamma_{\rm rad,\,T}$ needs to be maximized, which is equivalent, through equation (49), to maximizing the ratios $\gamma_{\rm max}/\gamma_{\rm rad,\,T}$ and $\gamma_{\rm max}/\gamma_{\rm cool}$. Hence, resolving the necessary scale hierarchy demands pushing the outer scales, $\gamma_{\rm cool}$ and $\gamma_{\rm max}$, as far away from each other as possible.

Let us now examine the implications this has on numerical cost. We begin by noting that $\gamma_{\rm cool}$ cannot be made arbitrarily small. Otherwise, the ambient upstream plasma becomes efficiently radiative, appreciably cooling down over the course of the simulation. Physically, this renders the initial background plasma temperature, T_0, meaningless and, hence, makes interpreting the simulation results more difficult; it also means, numerically, that the upstream Debye length, $\lambda_{\rm D,\,0}$, quickly becomes unresolved (we only resolve it marginally to begin with: $\lambda_{\rm D,\,0} = 1.2\Delta x$), leading to spurious numerical heating. If, to avoid this, one requires the upstream plasma cooling time to be at least some factor $M > 1$ longer than L/c, then equation (44) implies $\gamma_{\rm cool} = M\langle\gamma\rangle = 3M\theta_0$. Here, $\langle\gamma\rangle = 3\theta_0 = 72$ is the initial mean upstream Lorentz factor, and $t_{\rm cool,\,T}$ can be used instead of $t_{\rm cool,\,IC}$ because upstream particles cool in the Thomson regime. We find empirically that our simulations take $\simeq 3L/c$ for the reconnected magnetic flux to saturate, and thus we conservatively set $M = 6$, giving $\gamma_{\rm cool} = 430 = 3.6\times10^{-3}\sigma_{\rm c,0} = \gamma_{\rm max}/2500$.

What the preceding paragraph shows, importantly, is that, regardless of the particular choice for M, the minimum value of $\gamma_{\rm cool}$ is inevitably tied, through θ_0, to the Debye-length resolution requirement. The ratio $\gamma_{\rm max}/\gamma_{\rm cool}$ thus becomes a proxy for $L/\lambda_{\rm D,\,0} \simeq L/\Delta x \equiv N$, the number of cells (in the x-direction) across the simulation box. Specifically

$$N = L/\Delta x = 1.2L/\lambda_{\rm D,0} = 1.2L/\sqrt{\theta_0\sigma_{\rm c,0}}\,\rho_0 \simeq 12\gamma_{\rm max}/2\theta_0\sqrt{\sigma_{\rm h,0}} = 18M\gamma_{\rm max}/\gamma_{\rm cool}\sqrt{\sigma_{\rm h,0}} = 7680\,, \quad (52)$$

where we used the $\theta_0 \gg 1$ approximation, $\sigma_{\rm h,0} \simeq \sigma_{\rm c,0}/4\theta_0$. Our available resources limit us to $N = 7680$, and, hence, for $\sigma_{\rm h,0} = 1250$, to $\gamma_{\rm max}/\gamma_{\rm cool} \simeq 2500$. This ratio is, nevertheless, sufficient to realize a healthy Klein–Nishina energy-scale hierarchy.

With $\gamma_{\rm max}/\gamma_{\rm cool} = \sqrt{\gamma_{\rm max}/\gamma_{\rm rad,T}} = \sqrt{\gamma_{\rm rad,T}/\gamma_{\rm cool}}$ set, we must now decide where to place $\gamma_{\rm KN}$. We empirically find that a good fiducial choice is to set $\gamma_{\rm KN}/\gamma_{\rm rad,\,T}$ such that $\gamma_{\rm rad,\,IC} = \gamma_{\rm max}$. Raising $\gamma_{\rm KN}$ from here lowers $\gamma_{\rm rad,\,IC}$, somewhat inhibiting reconnection-powered NTPA and, hence, the production of high-energy gamma-rays and pairs. On the other hand, lowering $\gamma_{\rm KN}$ limits the pair-production optical depth, $\tau_{\gamma\gamma} \propto \gamma_{\rm KN}/\gamma_{\rm cool}$, and, simultaneously, the range of energies where particles are efficiently cooled, $t_{\rm cool,\,IC}(\gamma) < L/c$ (an extreme case of which are the $\gamma_{\rm KN}^{\rm (lo)}$ curves of Fig. 2). Overall, this slows the pair-production response of the system: high-energy particles take longer to radiate pair-producing photons, and those photons travel farther before being absorbed to create pairs. Choosing, then, $t_{\rm cool,\,IC}(\gamma_{\rm rad,\,IC}) = t_{\rm X}(\gamma_{\rm rad,\,IC}) = L/c$ to set $\gamma_{\rm rad,\,IC} = \gamma_{\rm max}$, gives $\gamma_{\rm KN} = 5500 = 0.046\sigma_{\rm c,0} = 0.26\gamma_{\rm rad,\,T} = 13\gamma_{\rm cool}$, and $\tau_{\gamma\gamma} = 3\gamma_{\rm KN}/5\gamma_{\rm cool} = 7.7$.

Although our fiducial $\gamma_{\rm KN}$ is finely tuned, this parameter is much more flexible in real astrophysical systems where numerical requirements do not limit the scale separation. With much larger values of $\gamma_{\rm max}/\gamma_{\rm cool}$, astrophysical reconnection may have $\gamma_{\rm KN}$ different from our numerical sweet spot while still preserving the important features: copious particle acceleration above $\gamma_{\rm KN}$, efficient cooling of high-energy particles, and short mean-free paths of emitted gamma-rays. A much more detailed discussion of this point for two concrete astrophysical systems – FSRQ jets and black hole accretion disc coronae – is presented by Mehlhaff et al. (2021).

Having set $\gamma_{\rm KN}$, $\gamma_{\rm cool}$, $\gamma_{\rm max}$, and $\tau_{\gamma\gamma}$, we are at last equipped to justify our choice of $\sigma_{\rm c,0}$. Even though $\sigma_{\rm c,0}$ is technically a non-radiative reconnection parameter (that we have already discussed in Section 3.1), it has unique consequences in the presence of Klein–Nishina radiative physics. Namely, $\sigma_{\rm c,0}$ needs to be at least a factor of several higher than $\gamma_{\rm KN}$ in order for the intrinsic X-point particle acceleration limit, $\gamma_{\rm X} = 4\sigma_{\rm c,0}$, to lie deeply in the regime where reconnection-energized particles emit pair-producing photons. As discussed by Mehlhaff et al. (2021), since the pair-production cross-section, $\sigma_{\gamma\gamma}$ [equation (14)], peaks at gamma-ray energies $\simeq 3.6(m_e c^2)^2/\epsilon_{\rm ph} \simeq 14\gamma_{\rm KN} m_e c^2$, and because a particle in the Klein–Nishina limit tends to donate about half its energy to a Comptonized photon, particles emitting photons at peak pair-production cross-section have typical Lorentz factors $\gamma \simeq 30\gamma_{\rm KN}$. To permit reconnection-energized particles to easily pass this limit, we have set $\sigma_{\rm c,0} = 22\gamma_{\rm KN}$ ($\gamma_{\rm X} = 88\gamma_{\rm KN}$). This is why our simulations feature such a high $\sigma_{\rm c,0}$ and, consequently, a relatively low $\gamma_{\rm max}/\sigma_{\rm c,0} = 0.1L/\sigma_{\rm c,0}\rho_0 = 9.1$ given their numerical size. The issue is that $\sigma_{\rm c,0}$ needs to be on the upper end of our energy-scale hierarchy, which is already strapped to maximize $\gamma_{\rm max}/\gamma_{\rm cool}$, and so $\sigma_{\rm c,0}$ ends up being somewhat close to $\gamma_{\rm max}$.

A summary of all our radiative and non-radiative simulation parameters is given in Table 1. In addition, we present a graphical description of how the various reconnection energy scales relate to key physical and numerical quantities in Fig. 3. The figure also illustrates how changing one quantity in this high-dimensional parameter space affects the others. For example, a simple system-size scan requires added care in the presence of all of these radiative effects, because changing L (e.g. by changing N), displaces not only $\gamma_{\rm max}$ higher, but also $\gamma_{\rm cool}$ lower, reducing the cooling time of the upstream particles with respect to L/c. Thus, to conduct such a scan, one would need to take care to set θ_0 low enough such that, even at the end of the scan (highest L), the upstream particles are still sufficiently cold that they do not radiate appreciably.

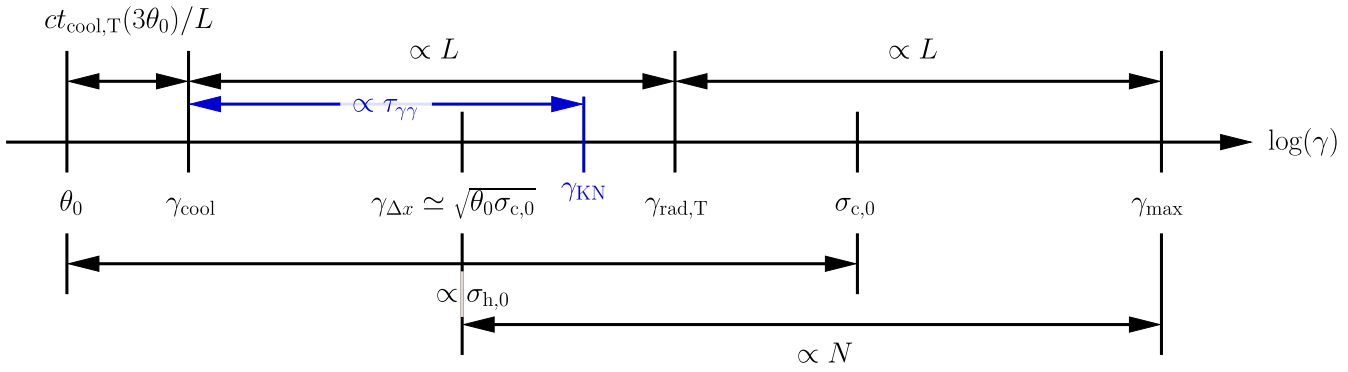

Figure 3. Important radiative ($\gamma_{\rm cool}$ and $\gamma_{\rm rad,\,T}$) and non-radiative (θ_0, $\sigma_{\rm c,0}$, and $\gamma_{\rm max}$) energy scales in reconnection connected with lines indicating how their relative magnitudes depend on: the system size, L; the number of simulation cells in the x-direction, N; the characteristic cooling time, $ct_{\rm cool,\,T}(3\theta_0)/L$, of the initial upstream particles (M in the text); and the hot magnetization, $\sigma_{\rm h,0}$. The labels $\sigma_{\rm h,0}$ and $ct_{\rm cool,\,T}(3\theta_0)/L$ follow only in the relativistically hot case, $\theta_0 \gg 1$. The Klein–Nishina Lorentz factor, $\gamma_{\rm KN}$, and, consequently, $\tau_{\gamma\gamma} \propto \gamma_{\rm KN}/\gamma_{\rm cool}$ [equation (51)], can be freely selected independently of all other diagrammed quantities (since $\gamma_{\rm KN}$ depends only on $\epsilon_{\rm ph}$, but not on any of the other parameters, like $U_{\rm ph}$ and L, that the other scales depend on). They are placed to illustrate the scale hierarchy of equation (50), $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, which is realized in all the simulations of this work that include Klein–Nishina effects.

4 KLEIN–NISHINA IMPACT ON RECONNECTION

Here, we present and compare the four main simulations whose setup is discussed in Section 3. We stress that, all parameters being the same (Table 1), the simulations differ only in their modelled physics. In our non-radiative run, 'no rad.', radiative cooling and pair production are completely turned off. In our Thomson-radiative run, 'IC(Th)', we apply continuous Thomson radiative cooling, but pair production remains absent. In the runs 'IC(KN)' and 'IC(KN)+PP', we employ our general IC cooling scheme (including the high-energy Klein–Nishina regime; Sections 2.2.1 and 2.2.3), but only self-consistently calculate pair production (Section 2.2.2) in the IC(KN)+PP run. In this sense, the IC(KN) run is artificial. In it, we simply pretend that all radiation emitted by particles, even the part above pair threshold, is permanently lost from the simulation.

In each of the following subsections, we analyse one aspect of the simulations, starting with those that are more similar among the four and moving to those that are more different. We finish by addressing unique properties of the run IC(KN)+PP that only exist in the context of pair production.

4.1 Global spatial evolution

We present the large-scale temporal evolution of the simulations in Fig. 4. For generality, the figure depicts the run, IC(KN)+PP, with general IC cooling and pair production, but the temporal evolution in terms of the spatial plasma number density (left-hand column) is similar in all runs.

At the simulation onset, we apply a 1 per cent perturbation to the in-plane magnetic field. This seeds the tearing instability in the initial current sheet, rapidly disrupting it into a chain of plasmoids separated by smaller current sheets, themselves tearing-unstable. The successive tearing of these current sheets down to smaller and smaller scales yields a self-similar hierarchy of plasmoids and inter-plasmoid current layers (Uzdensky, Loureiro & Schekochihin 2010), as in, for example, the snapshots of Fig. 4 at $t = 0.6L/c$ and $t = 1.2L/c$. As the plasmoids merge with one another, they also create separate miniature reconnection sites with current sheets oriented perpendicular to those of the main plasmoid chain (parallel to the yz-plane instead of to the xz-plane). These reconnection sites – for example at $x \simeq 35\sigma_{\rm c,0}\rho_0$ in the $t = 1.2L/c$ snapshot and at $x \simeq 5\sigma_{\rm c,0}\rho_0$ in the $t = 1.9L/c$ snapshot – beget their own plasmoid hierarchies.[4]

In both cases – whether in the mid-plane plasmoid chain or at vertical reconnection regions between colliding plasmoids – reconnecting current sheets are prominent sites of particle acceleration. This fact is underscored in the depicted IC(KN)+PP run by the copious emission of gamma-ray (above pair-production threshold) radiation from these regions in the middle column of Fig. 4.

Given the periodic boundary conditions of our setup, the large plasmoid with centre at $x \simeq 25\sigma_{\rm c,0}\rho_0$ in our simulations serves as an exhaust for the plasma processed by reconnection. Reconnection slows down and eventually stalls once about half of the initial magnetic flux in the box is reconnected. This corresponds to a state where the *separatrix* – the topological boundary between the domains of reconnected and unreconnected flux – crosses the mid-plane at an angle of about 45° (crosses itself at 90° angles) at the dominant X-point and opens up around the large exhaust plasmoid, which is then the only plasmoid remaining in the layer.

Finally, even though the runs are quite similar in their global spatial evolution, one unique aspect of the IC(KN)+PP run is the difference in spatial coherence between the original particles (Fig. 4, left-hand column) and those produced *in situ* (Fig. 4, right-hand column). Original particles are bound to magnetic field lines (except at reconnection X-points), and thus the striations in the original-particle number density follow closely the wrapping of magnetic field lines around large plasmoids. This forms a tree-ring pattern: density striations along field lines trace the history of magnetic flux accumulation onto each plasmoid – by merging with and absorbing smaller plasmoids – in the same way that rings on a tree stump trace the felled tree's lifecycle. In contrast, for the produced particles, such an effect, while still discernible, is much less pronounced. This owes to the added channel through which produced particles can take up residence in plasmoids: they can be born there directly. They are not constrained, like the original particles, to essentially follow the

[4]In a real instance of astrophysical reconnection, the outer scale, L, would dwarf the plasma microscales (e.g. $\sigma_{\rm c,0}\,\rho_0$) by many orders of magnitude, leading to a deep self-similar hierarchy both in the main plasmoid chain and in the recursive ones birthed between merging plasmoids. However, in our simulations with limited computational resources, we only witness the primary chain and the first secondary plasmoid-merger chains.

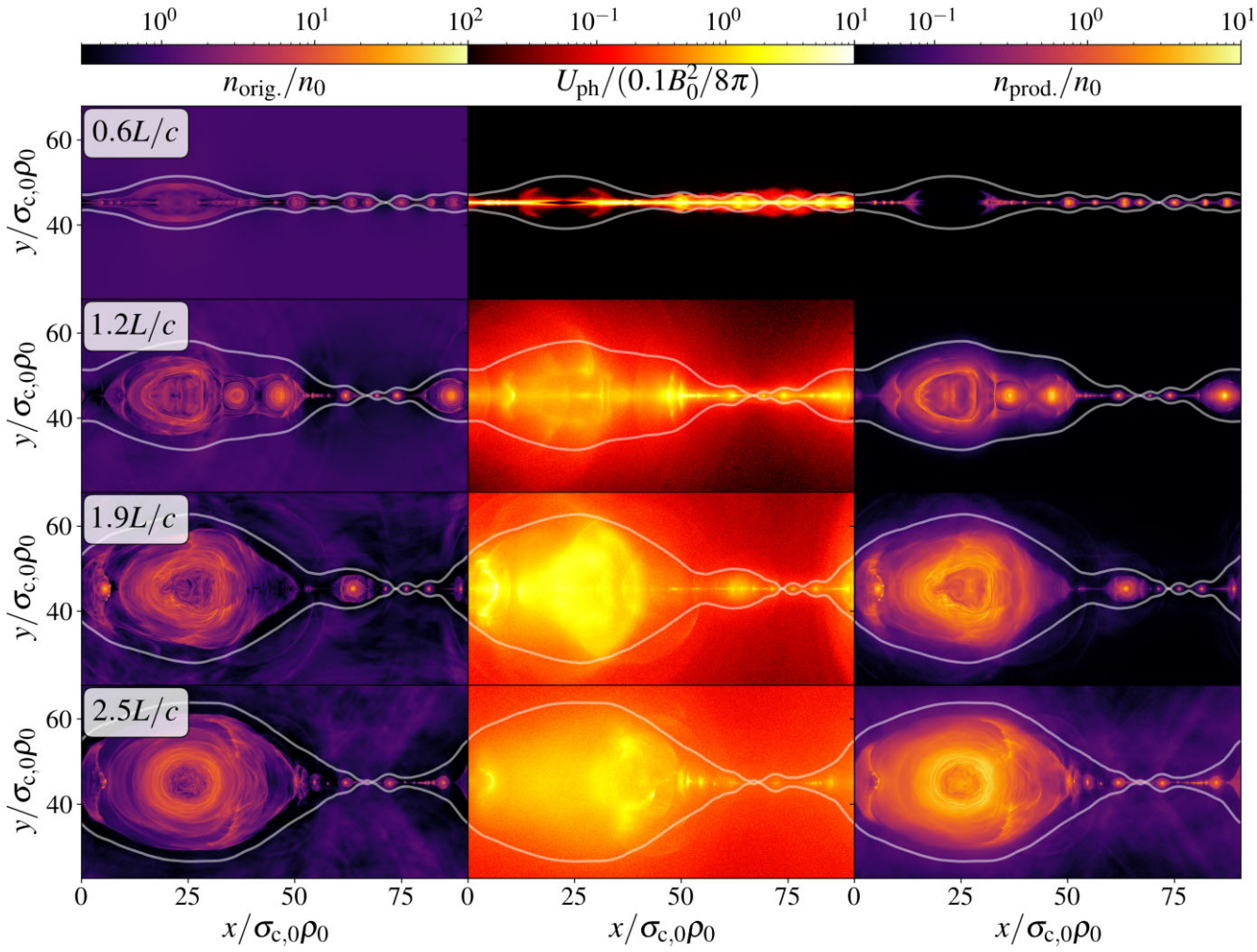

Figure 4. Time snapshots of the original pair number density (left-hand column), above-threshold photon energy density (middle column), and accumulated produced pair number density (right-hand column) for the main Klein–Nishina reconnection simulation with pair production, IC(KN)+PP. In each panel, the reconnection separatrix is drawn in white and intersects at the main reconnection X-point. All runs exhibit similar large-scale temporal evolution to that depicted: the initial current sheet quickly tears into a hierarchical chain of merging plasmoids (e.g. left-hand column). In addition to this familiar picture, in the IC(KN)+PP run, regions of active particle acceleration are signalled by flashes of gamma-rays (middle column). These photons are eventually absorbed to build up the population of particles born *in situ* (right-hand column).

reconnection of a given field line onto a plasmoid. This pollutes, in the number density maps of the produced particles, what are, on the maps of original-particle density, fairly pristine tree-ring-like density striations around plasmoids. These remarks are specific to 2D reconnection; in 3D, out-of-plane variation would likely obscure, even in the original particles, the tree-ring structures observed here.

4.2 Reconnection rate

Consistent with their similar global evolution, all four of our principal runs exhibit statistically indistinguishable reconnection rates, $\beta_{\rm rec}$. We measure these reconnection rates in Figs 5 and 6. Fig. 5 shows the reconnected flux, $\Phi(t)$, in each simulation as a function of time, t. To determine the characteristic reconnection rate, we consider the *active phase* of each simulation, defined as the period, $t \in [t_{\rm start}, t_{\rm end}]$, during which the middle 70 per cent of the change in reconnected flux (difference between initial and final points on the Fig. 5 curves) occurs. This insulates the measurement from artificially slow values during reconnection onset while also reducing sensitivity to the late-time slowdown during reconnection saturation (see discussion in Section 4.1). We define the average reconnection rate as $[\Phi(t_{\rm end}) - \Phi(t_{\rm start})]/B_0 c(t_{\rm end} - t_{\rm start})$. Our measurements are summarized in Fig. 6.

Even in the active phase, however, reconnection does not proceed at a precisely uniform rate. To provide some idea of the temporal variation, we also compute instantaneous reconnection rates by differentiating the curves in Fig. 5 with respect to time. We report the median and one standard-deviation (16th and 84th) percentiles of the resulting reconnection rate distributions in Fig. 6.

Concerning the differences between the simulations, the radiative runs generally reconnect about 20 per cent more magnetic flux than the non-radiative simulation. This effect was previously noted by Werner et al. (2019), who interpreted it as resulting from radiative cooling tending to reduce plasma pressure support inside plasmoids, enhancing plasmoid compression (cf. Hakobyan et al. 2019, 2023b; Schoeffler et al. 2019, 2023; Chernoglazov, Hakobyan & Philippov 2023). As a result, more reconnected magnetic flux is needed to achieve the saturated geometry where the separatrix forms 90° angles with itself at the principal X-point (Section 4.1). This effect primarily

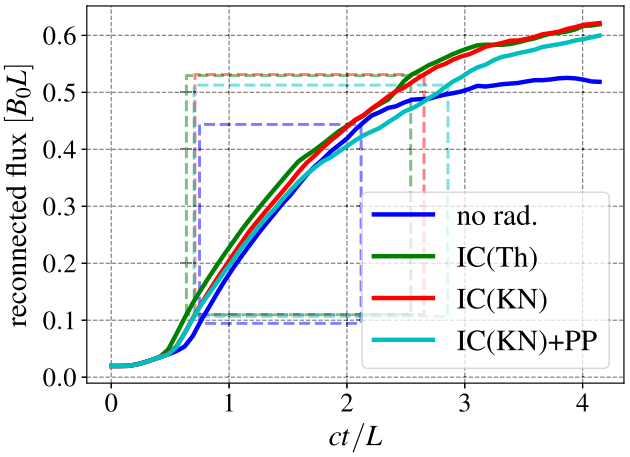

Figure 5. Reconnected magnetic flux as a function of time in our four main runs: no rad., IC(Th), IC(KN), and IC(KN)+PP. In this plot, we include information from both the top and bottom reconnection layers in our simulations. Thus, the saturation of each simulation's reconnected flux at $\sim 0.5 B_0 L$ indicates that the initially available flux in the box, $B_0 L_y/2 = B_0 L$, has been approximately half-consumed. The active phase of each run is displayed as a dashed box during which the middle 70 per cent of the simulation's total change in reconnected flux occurs. Each run's average reconnection rate (see Fig. 6) is the slope of the secant line through this box.

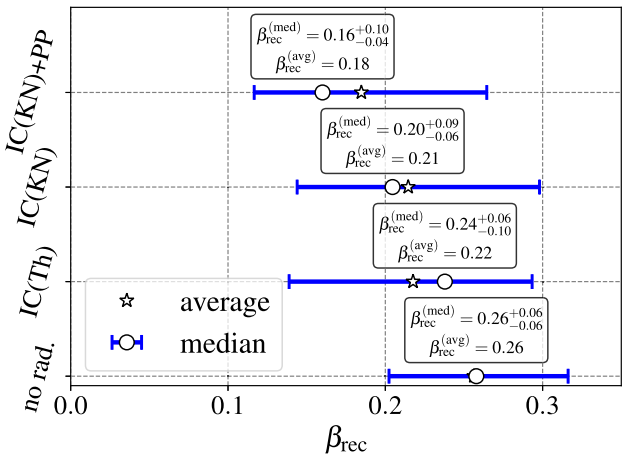

Figure 6. Per-simulation average, median, and 1σ (16th and 84th) percentile reconnection rates. The reconnection rates between runs are statistically indistinguishable.

alters the late-time flux saturation; it has little impact on the rate of reconnection during the active phases of our simulations.

Additionally, we observe some radiatively dependent skew in our measured reconnection rate distributions. Notably, the run IC(KN)+PP is almost bimodal in its reconnection rate distribution. As visible in Fig. 5, this run reconnects relatively quickly during the first half of its active period and slows down during the second half. As a result, the median reconnection rate falls substantially below the mean. This could be a sign of the pair feedback anticipated by Mehlhaff et al. (2021), where the pairs produced in the upstream region load the upstream plasma, thereby reducing $\sigma_{\rm h}$ and hence inhibiting reconnection (see also the supporting information of Xie et al. 2023), but it is not statistically significant. We examine the issue of pair feedback again in later sections.

4.3 Non-thermal particle acceleration

In our simulations, magnetic reconnection results in efficient energy delivery to the plasma particles in the form of non-thermal particle acceleration (NTPA). To illustrate this, we present the particle energy distributions, time-averaged over each simulation's reconnection active phase (Section 4.2), in Fig. 7. While NTPA is efficient in all cases, differences between the runs now begin to emerge. In particular, the slope and extent of the high-energy power-law tail differ depending on the radiative physics involved.

In the non-radiative run, reconnection promptly energizes particles up to Lorentz factors $\gamma \sim \gamma_{\rm X} \equiv 4\sigma_{\rm c,0}$ with a hard power law, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$, of index $p \simeq 1.2$. This is expected in our weak guide field, extremely highly magnetized ($\sigma_{\rm h,0} \sim 10^3$) regime, in which numerous previous studies have found that p tends to decrease (hardening the power law) with increasing $\sigma_{\rm h,0}$, asymptoting close to unity in the large-$\sigma_{\rm h,0}$ limit (e.g. Zenitani & Hoshino 2001, 2008; Guo et al. 2014, 2015, 2019, 2021; Melzani et al. 2014; Sironi & Spitkovsky 2014; Sironi et al. 2016; Werner et al. 2016, 2018; Werner & Uzdensky 2017; Ball, Sironi & Özel 2018; Sironi 2022; Uzdensky 2022). Because $p < 2$, most of the particle kinetic energy is stored in the high-energy tail of the distribution, forcing a departure from the $p \simeq 1.2$ scaling beyond Lorentz factors a few times $\gamma_{\rm X} \equiv 4\sigma_{\rm c,0}$. Otherwise, the particles would carry more energy than initially available in the reconnecting magnetic field.

Beyond the steepening around $\gamma = \gamma_{\rm X}$, the particle distribution declines and then eventually sharply cuts off near $\gamma = \gamma_{\rm max}$. Since the energies $\gamma_{\rm X} \equiv 4\sigma_{\rm c,0}$ and $\gamma_{\rm max}$ are relatively close together by computational necessity in this study (Section 3.2.2), it is difficult to determine whether the transition near $\gamma = \gamma_{\rm X}$ leads to a steeper (softer) power law, perhaps scaling roughly as $p \simeq 3$, or to an exponential cut-off that later gives way to an even sharper cut-off near $\gamma \simeq \gamma_{\rm max}$. If, of these two possibilities, it is the former case of a transition to a softer power law that is realized, it would match the picture advanced by Petropoulou & Sironi (2018) and Hakobyan et al. (2021) where particles undergo additional slow acceleration after being processed across the reconnection separatrix and becoming trapped inside of adiabatically compressing plasmoids.

The addition of radiative cooling markedly changes the signatures of NTPA in our simulations. In the purely Thomson radiative run, radiative losses impose a decisive cut-off on the maximum particle energy at the Lorentz factor $\gamma_{\rm rad,\,T}$, as first studied by Werner et al. (2019). This cut-off is well below not only the nominal system-size-limited Lorentz factor, $\gamma_{\rm max}$, but also the intrinsic maximum energy, $\gamma_{\rm X}$, attainable by particles via non-ideal direct acceleration by the reconnection electric field near X-points. Thus, radiative losses compete with even very rapid particle acceleration. Slower secondary acceleration channels are suppressed altogether, and there is no evidence of a secondary power-law component associated with such channels (consistent with earlier studies, e.g. Sironi & Beloborodov 2020; Mehlhaff et al. 2020, including when strong radiative cooling is facilitated by synchrotron losses instead of IC emission, as observed by Hakobyan et al. 2023b). The suppression of non-impulsive acceleration processes by (synchrotron or IC) radiative cooling has also been observed in the context of relativistic turbulence (e.g. Zhdankin et al. 2020; Comisso & Sironi 2021; Nättilä & Beloborodov 2021; Sobacchi, Nättilä & Sironi 2021; Zhdankin, Uzdensky & Kunz 2021). Had we conducted 3D simulations, the available secondary particle acceleration channels would have probably been faster than in the present 2D setup, with energized particles not indefinitely confined to plasmoids, but eventually escaping back into the upstream region to surf along the reconnection electric field on Speiser-like

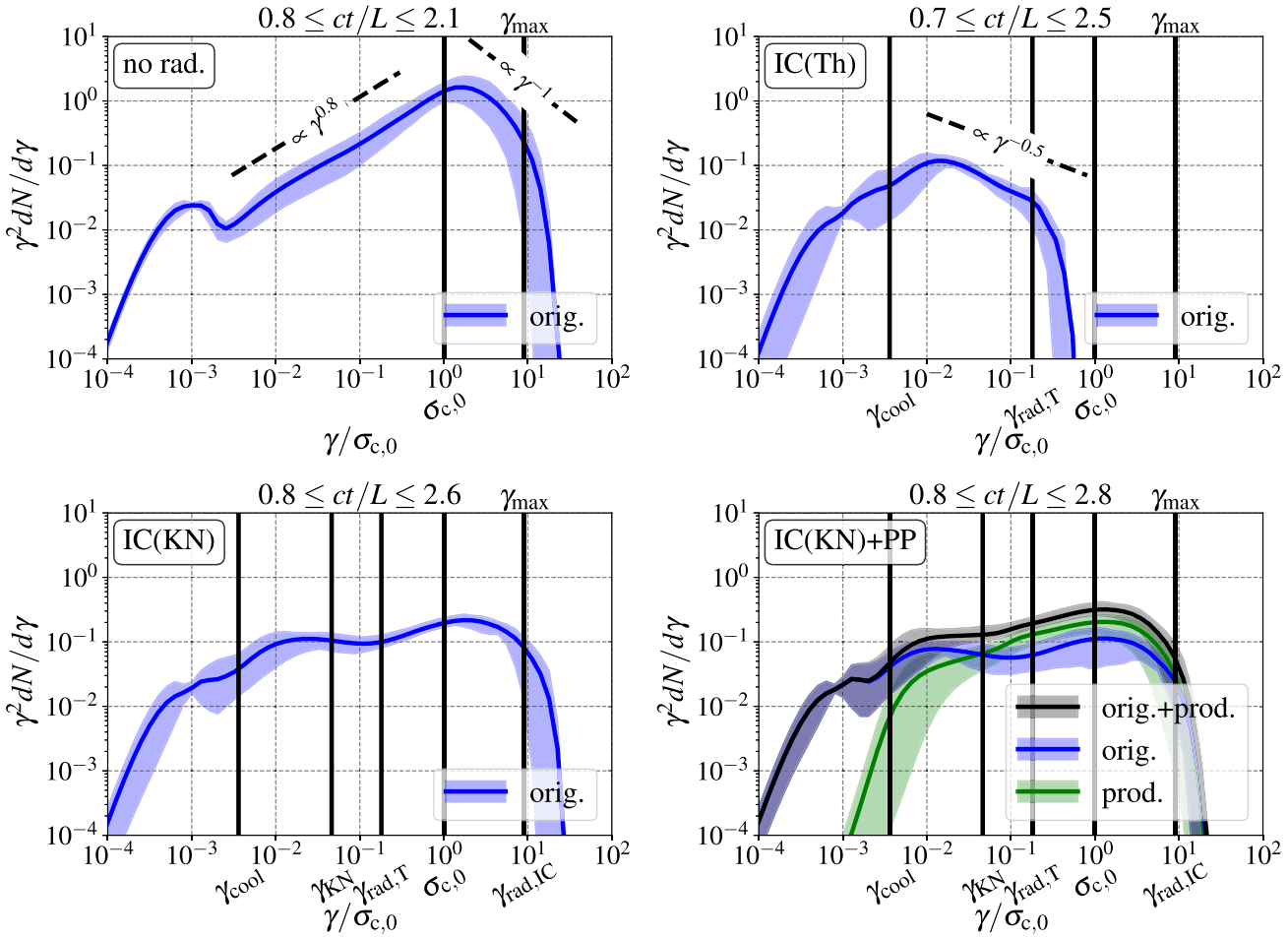

Figure 7. Time-averaged particle energy distributions for our four main runs. The averaging interval is the same as Fig. 5: the period during which the middle 70 per cent of the change in the reconnected flux occurs. Error envelopes display the 1σ (16th and 84th) percentiles during the averaging interval in each energy bin. The normalization is arbitrary, but equal on all panels. The IC(KN)+PP distribution is decomposed into contributions from particles that were originally present at time $t = 0$ (orig.) and those produced during the simulation (prod.). Klein–Nishina effects lead to NTPA that is intermediate between the non-radiative and Thomson radiative regimes, with a high-energy cut-off similar to the non-radiative case but a power-law scaling intermediate between those of the non-radiative and Thomson-cooled runs.

trajectories, rapidly accelerating to even higher energies (Zhang, Sironi & Giannios 2021; Zhang et al. 2023). Taking a hint from recent synchrotron radiative simulations (Chernoglazov et al. 2023), we expect that such acceleration would nevertheless remain suppressed in the presence of the intense radiative losses of the IC(Th) run, which would cool particles down before they could escape back into the upstream for further acceleration. These expectations must ultimately be checked by a future 3D study, though.

Besides the radiative cut-off at $\gamma_{\rm rad,\,T}$, the high-energy tail of the Thomson-cooled particle energy distribution features a steeper power-law scaling, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$, with $p \simeq 2.5$. This is also consistent with earlier work (Werner et al. 2019; Mehlhaff et al. 2020), including the pile-up seen at intermediate energies (yielding a peak in $\gamma^2\mathrm{d}N/\mathrm{d}\gamma$) just beyond $\gamma_{\rm cool}$. This pile-up results from intermittent episodes of explosive particle acceleration ignited at plasmoid mergers. In between such episodes, particles are rapidly cooled inside their host plasmoids, reaching a typical energy $\gamma_{\rm cool} L/c\,\Delta t_{\rm coll}$ characterized by the time $\Delta t_{\rm coll}$ between plasmoid collisions. While, in reality, $\Delta t_{\rm coll}$ is different for different tiers in the plasmoid hierarchy, the relativistic plasmoid motion in the box of size L dictates that it should be $< L/c$. This is consistent with the pile-up in the high-energy tail (peak in $\gamma^2\mathrm{d}N/\mathrm{d}\gamma$) occurring at a few-to-several times $\gamma_{\rm cool}$ in the IC(Th) panel (top right) of Fig. 7.

Klein–Nishina effects lead to an NTPA regime that is largely intermediate between the non-radiative and strongly Thomson-cooled cases. This is evidenced by both our runs with general Compton losses, IC(KN) and IC(KN)+PP. The Klein–Nishina reduction in radiative efficiency causes the particle energy distribution's tail to exhibit a flatter (harder) scaling, $p \simeq 2$, and to persist to higher energies (definitively cutting off by the time $\gamma \simeq \gamma_{\rm max} = \gamma_{\rm rad,\,IC}$, though perhaps steepening sooner, near $\gamma = \gamma_{\rm X}$) than when cooling proceeds purely in the Thomson limit. However, the tail is still steeper than that in the non-radiative run. This intermediate behaviour can be understood from the hierarchy of time-scales, $t_{\rm cool,\,T}(\gamma) < t_{\rm cool,\,IC}(\gamma) < L/c$, which holds at all Lorentz factors $\gamma < \gamma_{\rm max}$ in these simulations (Fig. 2, $\gamma_{\rm KN}^{\rm (med)}$ curves).

The degree to which NTPA more resembles that in the Thomson or non-radiative limits depends on the precise value of $\gamma_{\rm KN}$. Increasing $\gamma_{\rm KN}$ tends to bring the generalized cooling time, $t_{\rm cool,\,IC}(\gamma)$, closer to its Thomson limit, $t_{\rm cool,\,T}(\gamma)$ (Fig. 2, $\gamma_{\rm KN}^{\rm (hi)}$ curves). Once $\gamma_{\rm KN}$ surpasses $\gamma_{\rm rad,\,T}$, particles are forbidden from experiencing significant Klein–Nishina effects, and the system reverts to purely Thomson radiative

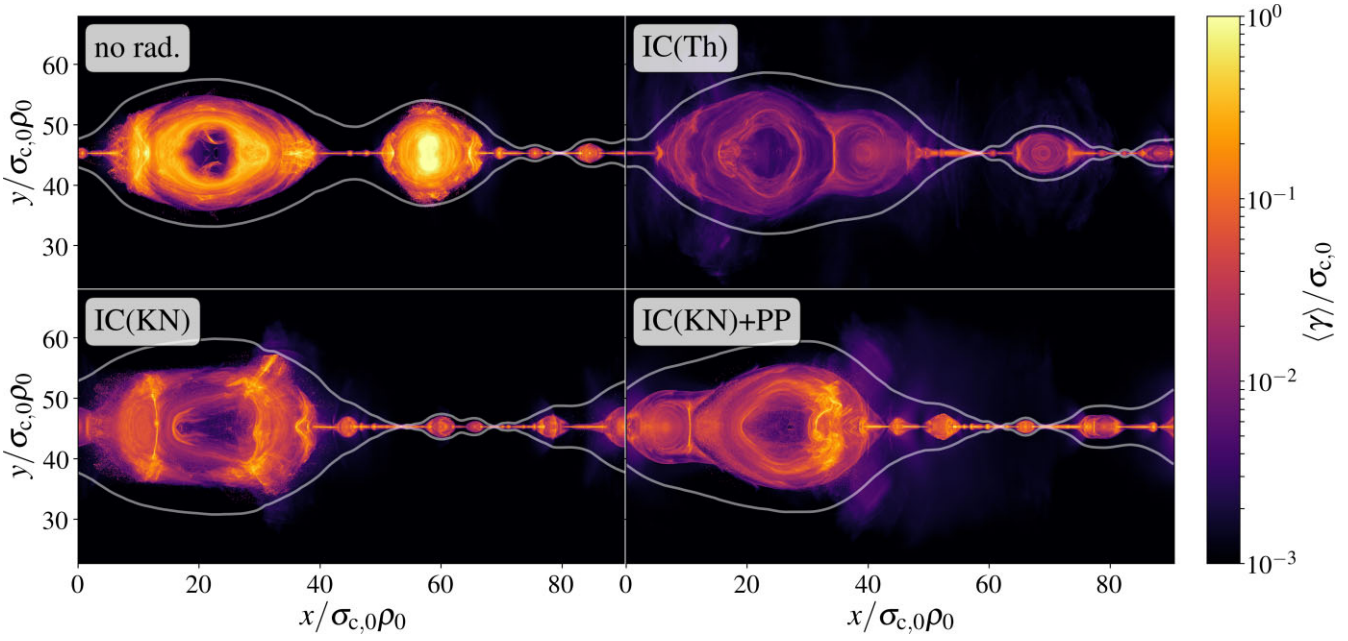

Figure 8. Local average Lorentz factor maps for our four main runs. Each run is pictured at the moment in time when half of its final change in reconnected magnetic flux has occurred. Commensurate with what happens in the particle energy distributions, Klein–Nishina effects result in maps that have properties that are intermediate between the non-radiative and purely Thomson-cooled regimes. Namely, the IC(KN) and IC(KN)+PP runs both feature hot reconnection current sheets like the non-radiative simulation, but plasmoids with temperatures between those of the plasmoids in the non-radiative and Thomson-cooled cases.

reconnection. On the other hand, reducing $\gamma_{\rm KN}$ lengthens the cooling time, $t_{\rm cool,\,IC}(\gamma)$, and, in the extreme case of $\gamma_{\rm KN} \lesssim \gamma_{\rm cool}$, $t_{\rm cool,\,IC}(\gamma)$ exceeds L/c for all γ: an effectively non-radiative regime (Fig. 2, $\gamma_{\rm KN}^{\rm (lo)}$ curves). We have verified these expectations by running simulations with differing $\gamma_{\rm KN}$ (those outlined in Table 2), though we do not present the NTPA from those simulations in detail. These runs confirm that increasing $\gamma_{\rm KN}$ moves the particle distribution cut-off closer to $\gamma_{\rm rad,\,T}$ while also steepening its power-law tail, causing NTPA to resemble the Thomson limit. Conversely, reducing $\gamma_{\rm KN}$ maintains the sharp particle energy cut-off at the system-size limit, $\gamma_{\rm max}$ (perhaps with an earlier cut-off or power-law transition near $\gamma_{\rm X}$), while simultaneously hardening the high-energy tail, transitioning the system towards non-radiative NTPA. Thus, the IC(Th) and no rad. runs represent the effective high- and low-$\gamma_{\rm KN}$ extremes, with the IC(KN) and IC(KN)+PP runs falling somewhere in the middle.

In Fig. 7, for the run with pair production, IC(KN)+PP, we decompose the particle energy distribution into its separate contributions from particles that were present originally in the simulation and those produced on the fly. Strikingly, the produced particles dominate not only the high-energy tail, but also the total plasma energy contained in the simulation box. This is in spite of the fact that these particles are far less numerous than their originally present counterparts, which is only possible thanks to their extremely high average energy. That the produced particles should compete with the original particles for energetic dominance despite being fewer in number is in line with the basic predictions for this regime of reconnection advanced by Mehlhaff et al. (2021). We examine more thoroughly the differences between original and produced particles in Section 4.6.

To complement the distributions of Fig. 7 with a spatial view into particle acceleration, we present maps of the local plasma average Lorentz factor for each of our four main runs in Fig. 8. Here again, we see that the non-radiative and Thomson-cooled cases represent two opposite extremes. On the one hand, the non-radiative run contains hot reconnection current sheets and plasmoids (except the cold core of the large exhaust plasmoid centred at $x \simeq 20\sigma_{c,0}\,\rho_0$ – it is composed of initially drifting plasma swept directly into this plasmoid near the beginning of the simulation without ever being energized by reconnection). On the other hand, the simulation with Thomson losses features cold plasmoid cores (cf. Beloborodov 2017; Sironi & Beloborodov 2020; Sridhar et al. 2021, 2023; Groselj et al. 2023) and merely warm current sheets.

Let us see how these differences arise. The plasmoids in the non-radiative run accumulate kinetic energy via the hot plasma that is exhausted away from reconnection X-points, thereby containing, collectively, a running tally of the dissipated magnetic energy. While plasmoids still collect particles in the IC(Th) case, they no longer amass liberated energy, which instead escapes as IC radiation. Then, instead of the area-filling, particle-accumulating plasmoids, it is the quasi-1D current-sheet singularities that host the energetic plasma – that is, where intense acceleration is actively taking place.[5] And even in these special regions, the local mean energy is radiatively limited to $\sim\gamma_{\rm rad,\,T} = 0.2\sigma_{c,0}$, much lower than the intrinsic maximum X-point acceleration Lorentz factor, $\gamma_{\rm X} = 4\sigma_{c,0}$, reached in the (consequently much hotter) current sheets of the non-radiative run. In the IC(Th) case, once particles vacate rapid acceleration zones near X-points to move into plasmoids, they quickly cool down, giving the plasmoid cores a characteristic mean energy of order the pile-up energy, $\langle\gamma\rangle \sim 10^{-2}\sigma_{c,0}$, in the IC(Th) distribution of Fig. 7.

Moving now to the IC(KN) and IC(KN)+PP simulations in Fig. 8, we see again that they are intermediate between the non-radiative and Thomson-cooled extremes. Like the non-radiative case, these two runs contain very hot reconnection current layers, with local Lorentz factors comparable to $\sigma_{c,0}$ and far exceeding $\gamma_{\rm rad,\,T}$. This reflects

[5] In 3D, the plasmoids would be volume-filling instead of area-filling, and the current sheets quasi-2D structures instead of quasi-1D ones.

the fact that cooling losses do not substantially inhibit acceleration near reconnection X-points in these Klein–Nishina-regime runs. As for plasmoids, these are colder than in the non-radiative simulation but warmer than in the IC(Th) case. This stems again from the cooling time-scale hierarchy, $t_{\mathrm{cool,\,T}}(\gamma) < t_{\mathrm{cool,\,IC}}(\gamma) < L/c$: particles accelerated at current sheets are not efficiently cooled on the time-scales of their acceleration, but they are still efficiently cooled over one dynamical time, causing plasmoids to cool down – just not as quickly as in the IC(Th) run.

To summarize, in this subsection, we have witnessed the first main differences emerge among our simulations, with the differing radiative physics leaving pronounced and distinguishing imprints on NTPA. Even though all runs exhibit non-thermal power-law particle energy distributions, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$, in their reconnection active phases, the slopes and extents of their power-laws differ dramatically. The non-radiative regime yields p close to unity, with a departure from this scaling near $\gamma = \gamma_{\mathrm{X}} \equiv 4\sigma_{\mathrm{c},0}$ and a subsequent sharp cut-off at $\gamma = \gamma_{\mathrm{max}}$. The IC(Th) run represents an opposite regime, with $p \simeq 2.5$ followed by an abrupt cut-off at $\gamma \simeq \gamma_{\mathrm{rad,\,T}} \ll \gamma_{\mathrm{max}}$. Klein–Nishina radiative cooling lies between these two extremes, and for the parameters of our IC(KN) and IC(KN)+PP runs, gives $p \simeq 2$ and a sharp cut-off at $\gamma \simeq \gamma_{\mathrm{max}}$ (perhaps with an earlier steepening near $\gamma = \gamma_{\mathrm{X}}$).

4.4 Inverse Compton emission spectra

In this section, we connect the pronounced NTPA in our simulations to their IC emission spectra. Like the underlying particle energy distributions (Section 4.3), these spectra are highly extended and non-thermal, and the distinctions among them reflect the differing radiative physics at play.

We present angle-integrated IC emission spectra for each of our four main runs in Fig. 9. These spectra are computed by summing together the individual spectra from every simulated particle, regardless of the particle's position or velocity direction. As a result, beaming and light-traveltime effects are neglected. In addition, for the IC(KN)+PP run, pair-production absorption of the emitted gamma-rays is temporarily ignored, but we discuss it briefly at the end of this subsection and in more detail in the next subsection.

In the non-radiative and Thomson-cooled cases, the IC emission spectra are produced in the Thomson regime and are thus given straightforwardly in terms of the underlying particle energy distributions as follows. Particles of Lorentz factor γ upscatter photons to energies $\epsilon(\gamma) \sim \gamma^2\epsilon_{\mathrm{ph}}$. Because the scattering rate, $\sigma_{\mathrm{T}}cn_{\mathrm{ph}}$ [cf. equation (5)], is independent of γ, the number of photons emitted per unit time into a given energy interval, $\mathrm{d}\epsilon\mathrm{d}N_{\mathrm{ph}}/\mathrm{d}t\mathrm{d}\epsilon$, is proportional to the number of particles at the corresponding scattering Lorentz factor, $\mathrm{d}\gamma\mathrm{d}N/\mathrm{d}\gamma|_{\gamma\propto\epsilon^{1/2}}$. Thus, if the particle energy distribution is a power law, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$, the emitted photon distribution is also a power law: $\mathrm{d}N_{\mathrm{ph}}/\mathrm{d}t\mathrm{d}\epsilon \propto (\mathrm{d}N/\mathrm{d}\gamma|_{\gamma\propto\epsilon^{1/2}})(\mathrm{d}\gamma/\mathrm{d}\epsilon) \propto \epsilon^{-(p+1)/2}$ (cf. Rybicki & Lightman 1979). In the $\epsilon F_{\mathrm{IC}}(\epsilon)$ representation plotted in Fig. 9, this translates to $\epsilon F_{\mathrm{IC}}(\epsilon) = \epsilon^2\mathrm{d}N_{\mathrm{ph}}/\mathrm{d}t\mathrm{d}\epsilon \propto \epsilon^{-(p\,-\,3)/2}$.

This result equips us to easily interpret the power-law components in the non-radiative and Thomson-cooled IC emission spectra. The non-radiative run's particle distribution power law is approximately $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-1.2}$ (Fig. 7), which yields the expected IC power law, $\epsilon F_{\mathrm{IC}}(\epsilon) \propto \epsilon^{0.9}$. This is in good agreement with the non-radiative IC spectrum in the top left panel of Fig. 9. Additionally, if present, the second/steeper particle power-law component for this same simulation, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-3}$ – putatively stemming from slower secondary acceleration channels – should produce a flat spectrum, $\epsilon F_{\mathrm{IC}}(\epsilon) \propto \epsilon^0 = \mathrm{const}$. While this is roughly consistent with the measured spectrum of Fig. 9, it is difficult to definitively say that such a component truly exists and is not just part of the spectral cut-off. Regarding the Thomson-cooled run, the power law, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-2.5}$, should yield the gently increasing spectrum, $\epsilon F_{\mathrm{IC}}(\epsilon) \propto \epsilon^{0.25}$, which agrees with Fig. 9 (top right panel). Finally, all these ideas can be applied not just to connect spectral slopes between the particle and IC spectra, but also to relate their cut-offs. In particular, the cut-offs at γ_{max} and $\gamma_{\mathrm{rad,\,T}}$ in the respective non-radiative and Thomson-cooled particle distributions (Fig. 7) correspond to the observed cut-offs at photon energies of order $\gamma_{\mathrm{max}}^2\epsilon_{\mathrm{ph}}$ and $\gamma_{\mathrm{rad,T}}^2\epsilon_{\mathrm{ph}}$ in these simulations' IC emission.

This simple framework breaks down in the presence of Klein–Nishina effects. Then, the characteristic scattered photon energy becomes a broken power-law function of the particle's energy: $\epsilon(\gamma) \sim \gamma^2\epsilon_{\mathrm{ph}}$ when $\gamma < \gamma_{\mathrm{KN}}$ and $\epsilon(\gamma) \sim \gamma m_e c^2$ otherwise. Furthermore, the scattering rate, $\sigma_{\mathrm{T}}cn_{\mathrm{ph}}g_{\mathrm{KN}}(\gamma/\gamma_{\mathrm{KN}})$ [equation (5)], becomes a non-trivial, decreasing function of γ. This suppresses the emission efficiency and breaks the simple correspondence between the particle distribution power-law index and that of the IC emission spectrum. Let us examine how these effects manifest themselves in the IC(KN) and IC(KN)+PP panels (bottom left and bottom right, respectively) of Fig. 9. To begin with, the characteristic emission energies from particles at each of our Lorentz-factor scales are pushed closer together beyond the energy $\epsilon(\gamma_{\mathrm{KN}})$ (because, beyond γ_{KN}, ϵ scales linearly with γ instead of quadratically). As an example, even though the scales γ_{max}, $\gamma_{\mathrm{rad,\,T}}$, and γ_{cool} are all equally spaced on a logarithmic scale (because $\gamma_{\mathrm{rad,T}}^2 = \gamma_{\mathrm{max}}\gamma_{\mathrm{cool}}$), the corresponding photon energies, $\epsilon(\gamma_{\mathrm{max}})$, $\epsilon(\gamma_{\mathrm{rad,\,T}})$, and $\epsilon(\gamma_{\mathrm{cool}})$, are not at all evenly spaced, with $\epsilon(\gamma_{\mathrm{max}})$ and $\epsilon(\gamma_{\mathrm{rad,\,T}})$ closer together than $\epsilon(\gamma_{\mathrm{rad,\,T}})$ and $\epsilon(\gamma_{\mathrm{cool}})$ [because the former two lie above the break energy $\epsilon(\gamma_{\mathrm{KN}})$ while the latter lies below it]. Next, even though the power-law scalings of the IC(KN) and IC(KN)+PP particle distributions are both approximately γ^{-2} or shallower (i.e. harder), the corresponding $\epsilon F_{\mathrm{IC}}(\epsilon)$ spectra are both steeper (i.e. softer) than the Thomson-limit prediction, $\epsilon^{0.5}$, demonstrating the reduced radiative efficiency in the Klein–Nishina limit.

While the diminished Klein–Nishina cross-section produces a softer emission spectrum for a given particle distribution, it also yields a particle distribution that is harder in the first place (Section 4.3). These two effects somewhat cancel out, and, hence, not much change is observed in the spectral slope from the time-averaged IC(Th) spectrum to those yielded by the IC(KN) and IC(KN)+PP simulations (cf. Moderski et al. 2005). This is even despite the very different shape – the result of different cooling physics – in the particle distributions between these runs.

Finally, we note that the IC(KN) and IC(KN)+PP Compton emission spectra peak far above $\epsilon(\gamma_{\mathrm{KN}})$ – deeply in the Klein–Nishina regime. As a result, most of the radiated energy in the IC(KN)+PP case (grey region in the lower right panel of Fig. 9) is emitted above pair-production threshold, where [unlike the IC(KN) run] it will be recaptured by the system as hot newborn pairs. The peak in the intrinsic emitted IC spectrum is therefore invisible to the observer, who sees only the indirect remnant of this radiation reprocessed to below-threshold energies. We elaborate the observable consequences of this effect in the next subsection.

4.5 Light curves and spectral variability

We now complement Section 4.4's energy-resolved view of the emission from our simulations by discussing the timing of the radiative signatures. When viewed through the lens of timing, the

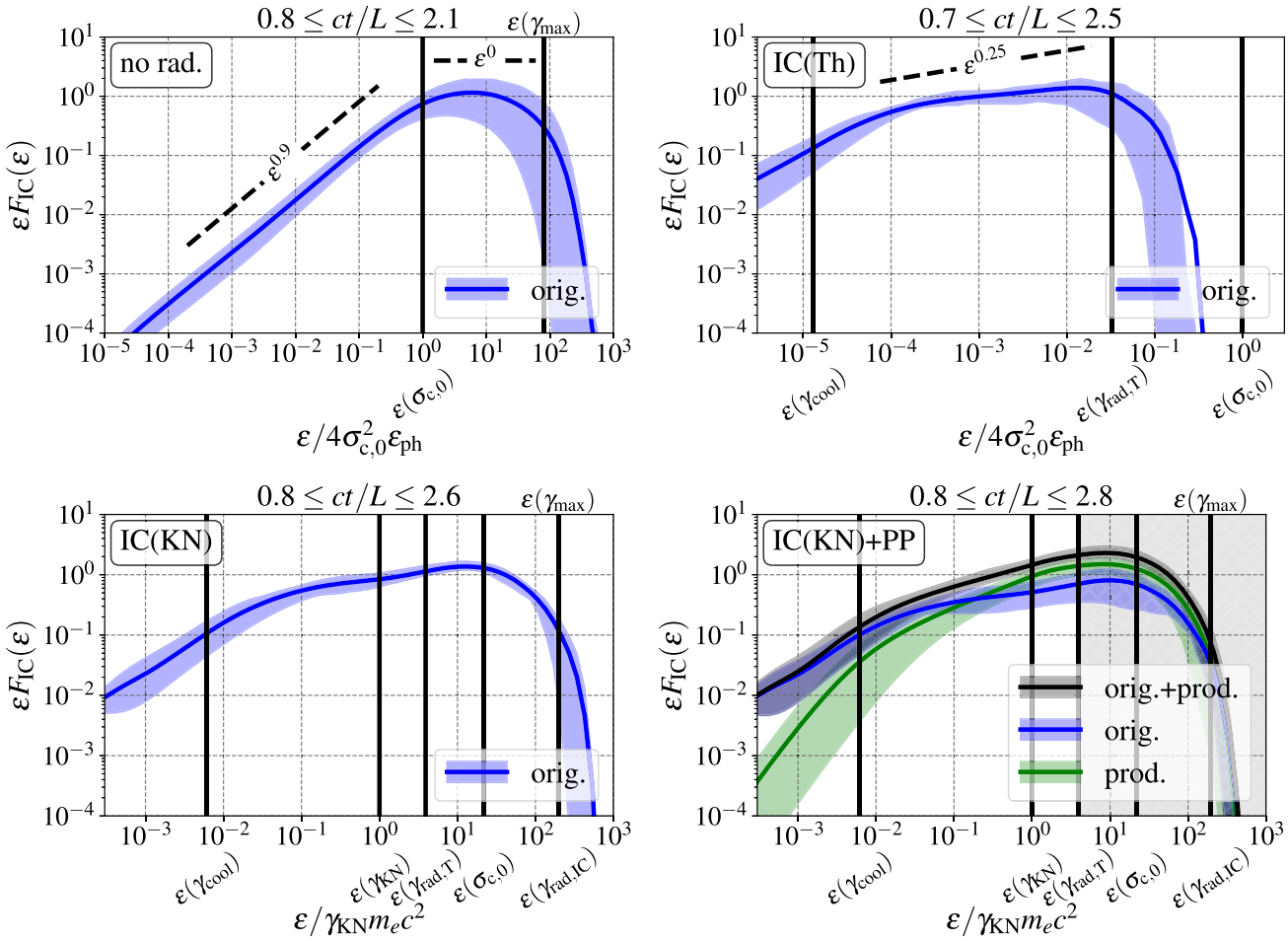

Figure 9. Time-averaged (over the same interval as in Figs 5–7 IC emission spectra for our four main runs. Instantaneous spectra used for averaging are equal to the box-integrated isotropic IC emissivity, $\epsilon F_{\rm IC}(\epsilon;t) \equiv \epsilon \int {\rm d}^3x {\rm d}\Omega\, j_{\rm IC}(\epsilon, \boldsymbol{x}, \Omega; t)$. Error envelopes display the 1σ percentiles over the averaging interval in each energy bin. The normalization is arbitrary but equal across all panels. For the no rad. run, besides the normalization, the luminosity scale itself is arbitrary – not tied to the energetics of the simulation – and is chosen so that the spectral peak is at a similar level to those of the other runs. The characteristic photon energy, $\epsilon(\gamma)$, emitted by a particle of Lorentz factor γ is defined piecewise continuously as $4\gamma^2\epsilon_{\rm ph}$ (γm_ec^2) for γ less (greater) than $\gamma_{\rm KN}$. The IC(KN)+PP spectrum is decomposed into contributions from originally present (orig.) and produced (prod.) particles. Klein–Nishina effects suppress the IC cross-section, hardening the particle energy distribution (Fig. 7), while softening the emitted spectrum. These competing effects result in IC(KN) spectral scalings that are not very different from that in the IC(Th) simulation. The grey region in the IC(KN)+PP panel indicates photon energies beyond pair-production threshold, $4\gamma_{\rm KN}m_ec^2$. Photons emitted at these energies do not make it to the observer; they are absorbed inside the system to produce new pairs.

differences among the various radiative regimes are accentuated, resulting in highly distinct observable signatures.

We begin by presenting light curves of each run's bolometric luminosity (instantaneous total escaping emitted power) in Fig. 10. For consistency with the flow of energy in the simulations, we only include that part of the luminosity permanently lost by the simulation. This means that, for the light curves of the no rad. and IC(Th) runs, we report the frequency-integrated emitted power as a function of time. In contrast, for the IC(KN)+PP run, we report only the portion of the emission spectrum below pair-production threshold: at photon energies $\epsilon < 4\gamma_{\rm KN}m_ec^2$. By the same reasoning, we are obligated to include all photon energies for the IC(KN) simulation, reporting the total emitted power in that case also – otherwise, since pair production is artificially suppressed in that run, we would not fairly count the energy lost from the simulation. We have checked that this bookkeeping yields the same total energy radiated (integrals of the curves in Fig. 10) by each radiative simulation.

The light curves in Fig. 10 are clearly separated into two main groups: the non-radiative versus the radiative simulations. This dichotomy excellently illustrates a fundamental property of radiative reconnection: prompt emission. That is, in non-radiative reconnection, particles are first accelerated and, then, over much longer timescales than the duration of the reconnection process itself, radiate away their energy as potentially observable emission. In contrast, radiative reconnection features fundamentally prompt emission, where particles radiate their acquired energy on subdynamical timescales, causing radiation to participate in the reconnection dynamics rather than, as in the non-radiative case, passively trace energization that has already occurred. This is reflected in Fig. 10 in that the three radiative light curves broadly track the instantaneous electromagnetic dissipation of their simulations, rising as reconnection gets going and falling again once the reconnected flux saturates (cf. Fig. 5). The light curve of the non-radiative simulation, on the other hand, follows the time integral of the electromagnetic dissipation, growing with the cumulative dissipated energy and reconnected flux.

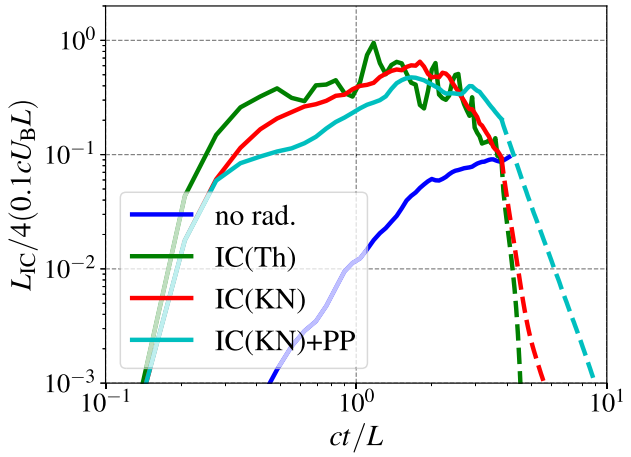

Figure 10. Light curves of the IC luminosity (instantaneous escaping radiated power) for each of our four main runs. The normalization for the non-radiative simulation is not fixed by dynamical radiative cooling and so is scaled to appear on the plot with the other simulations (cf. Fig. 9). Dashed portions of light curves are computed in post-processing (see the text).

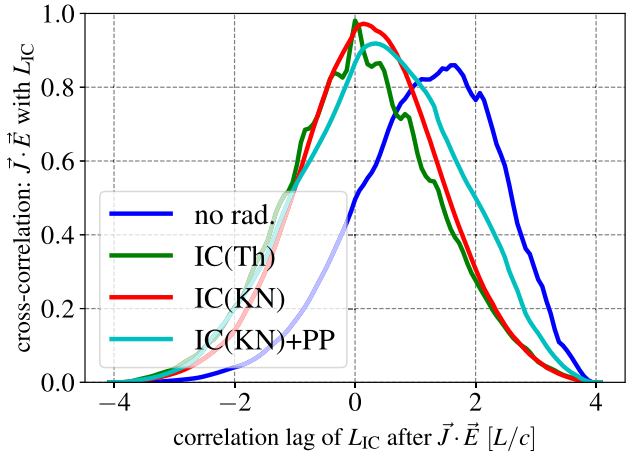

Figure 11. Cross-correlation of the instantaneous rate of electromagnetic dissipation with the escaping IC luminosity (light curves of Fig. 10). Emblematic of the distinction between radiative and non-radiative reconnection, the non-radiative simulation's luminosity lags its dissipation (proportional to the cumulative dissipated electromagnetic energy), whereas the radiative simulations' luminosities are synchronized with dissipation (proportional to its instantaneous rate). The IC(KN) run features a slightly longer lag than the IC(Th) run owing to its lower radiative efficiency. The IC(KN)+PP run features a still somewhat longer lag with a skewed distribution toward the high-lag end. This stems from the time it takes for radiation injected above pair-production threshold to become reprocessed to below-threshold (and, hence, escaping) energies.

To illustrate these remarks more thoroughly, we also supply Fig. 11, which shows the cross-correlation function, $\propto \int {\rm d}t\, L_{\rm IC}(t+\tau)P_{\rm diss}(t)$, with τ the correlation lag, of each $L_{\rm IC}(t)$ curve from Fig. 10 with the box-integrated electromagnetic dissipation, $P_{\rm diss}(t) \equiv \int {\rm d}^3x\, \boldsymbol{J}(t)\cdot\boldsymbol{E}(t)$. As discussed above and confirmed by this figure, there is a significant lag, $\tau > 0$, of the non-radiative IC luminosity behind $P_{\rm diss}$. In contrast, all three radiative simulations lose their energy promptly, with peak lags close to zero.

There are, however, differences among the radiative runs. The IC(Th) luminosity exhibits the shortest variability time-scales in Fig. 10, with small bumps atop its overall envelope corresponding to bursts of particle acceleration at plasmoid mergers. In contrast, because the cooling time-scales are slightly longer for the IC(KN) and IC(KN)+PP runs, particles retain enough energy in between these episodic events to smooth out the variations in the resulting light curves. As one might expect from these remarks, the IC(Th) simulation exhibits the smallest (exactly zero) peak lag from its electromagnetic dissipation to its emitted luminosity. In contrast, the IC(KN) and IC(KN)+PP runs have small but finite peak lag. Of these, the IC(KN)+PP case has a slightly longer lag and a more skewed cross-correlation distribution, with more power concentrated at longer lags. This is the result of the reprocessing of gamma-rays (near the peak of the emission spectrum in Fig. 9) to lower energies through gamma-ray radiation and pair production. That is, it takes time for power injected at the high-energy, above-threshold peak of the IC emission spectrum to be processed down to lower energies where it can escape the system.

When combined with spectral information, the lag induced by pair-production and gamma-ray absorption described above leaves a prominent imprint on the observable signatures of Klein–Nishina radiative reconnection – one that could potentially distinguish it from other radiative regimes. To demonstrate this, we present a side-by-side comparison of the luminosities and average photon energies, $\langle\epsilon\rangle \equiv \int {\rm d}\epsilon\, \epsilon F_{\rm IC}(\epsilon) / \int {\rm d}\epsilon\, F_{\rm IC}(\epsilon)$, viewed as a function of time for the IC(Th) and IC(KN)+PP runs in Fig. 12.

To fully appreciate the observational differences between these runs, we extrapolate their spectra beyond the end of each simulation. This is done under the assumption that the simulations are evolved to the point where no further exchange of energy occurs between the particles and the fields. Hence, the particle energy distributions can be passively Compton cooled in 1D energy space. Pair production can be included in this 1D evolution thanks to the homogeneous, isotropic, and static nature of the seed photon bath, which introduces no spatial or velocity-direction dependence. Specifically, the post-processing is implemented by sampling the particle and photon energy distributions near the end of the simulation (at a time-step after the reconnection active period; cf. Section 4.2) with a large number of quasi-particles and quasi-photons: like a PIC code, except that the quasi-particles and quasi-photons only have one coordinate – their energy – rather than six corresponding to their position and momentum vectors. Then, particle cooling is employed after the method described in Section 2.2.3, while pair production is calculated using the procedure from Section 2.2.2. Post-processing, in this way, the luminosity and average photon energy time-series allows us to extend them in Fig. 12 from just shy of $4L/c$ to more than $10L/c$ – a significant gain over the simulations themselves, which would otherwise need to be expensively evolved to more than double their actual duration to reach the same times.

Several facts, each of them accessible to observations, are immediately apparent from Fig. 12. First, the overall luminosity and average photon energy emitted from the IC(Th) simulation are tightly correlated. During the short time-scale variations of the light curve, sporadic magnetic reconnection acceleration yields an extended non-thermal distribution of particles (Fig. 7) with the resulting IC emission spectrum peaking near the high-energy cut-off (Fig. 9): that is, the acceleration mechanism produces a correlation between $\langle\epsilon\rangle$ and $L_{\rm IC}$ on the rising side of each subpeak in their time-series. In between these reconnection energization episodes, particles emitting at the highest energies – the ones controlling both the overall luminosity and peak photon energy – are also the most rapidly cooled. They thus suddenly plummet to lower energies, inducing

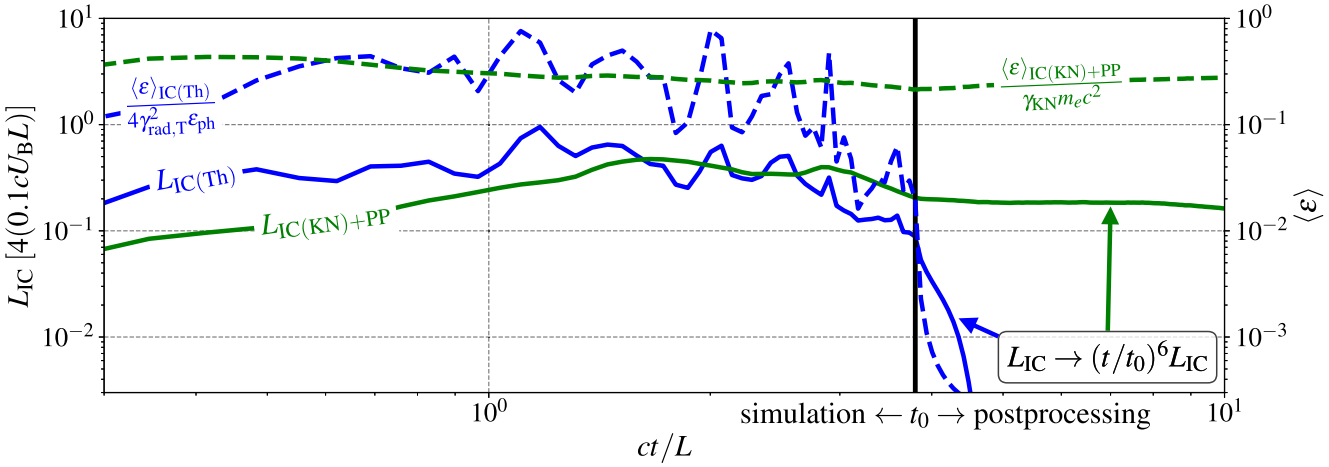

Figure 12. Light curves (reproduced from Fig. 10) of escaping IC luminosity and overall (computed using the escaping spectrum) average photon energy for the IC(Th) and IC(KN)+PP simulations. Both quantities are extrapolated beyond the simulation (from time t_0 indicated by the vertical black bar) via post-processing (see the text). After t_0 (but not before), the luminosities are compensated by $(t/t_0)^6$ – an empirical choice. The time-series of mean photon energy are never compensated. Several robust trends are evident that could distinguish these two types of reconnection in the context of reconnection-powered flares. Thomson radiative reconnection features a tight correlation between IC luminosity and mean photon energy. In contrast, the Klein–Nishina reconnection mean photon energy is $\sim 0.3\gamma_{\rm KN}m_ec^2$ independently of luminosity. These patterns hold both in the reconnection active phase and in the decaying phase when energization from the fields is either slowing down (as at late simulation times) or completely shut off (as in the post-processing zone of the plot). In the limit of no further energy injection, the t^{-6} brightness decay visible in the Klein–Nishina run sharply contrasts the much more precipitous decline in the Thomson case.

simultaneous drops in both $\langle\epsilon\rangle$ and $L_{\rm IC}$: that is, the radiative cooling mediates the correlation between $\langle\epsilon\rangle$ and $L_{\rm IC}$ on the falling side of each peak in their time-series.

A second observationally pertinent property of the IC(Th) curves in Fig. 12 is that, once energization from the electromagnetic fields is shut off, both the IC photon energy and luminosity drop precipitously. In fact, even though the luminosity time-series are compensated by $(t/t_0)^6$ rightward of the transition time, t_0, to the post-processing stage, this does little to stem the fall of the IC(Th) luminosity. Flares from highly radiative Thomson reconnection are thus characterized by *tight correlation between the observed luminosity and photon energy plus a rapid falling phase where both plummet together*.

Let us now examine how these qualities compare to those of the IC(KN)+PP run. Most importantly, the tight correlation between luminosity and photon energy is broken. Instead, irrespective of the instantaneous IC brightness, the average photon energy remains rock steady, persisting near $0.3\gamma_{\rm KN}m_ec^2$ – even in the decaying phase of the light curve after electromagnetic energization has ceased. In the following, we argue that this average photon energy is first set during active periods of reconnection-powered particle acceleration and subsequently reinforced, when such acceleration is inactive, by the Klein–Nishina radiative physics, explaining its persistence.

During active acceleration episodes, reconnection produces a hard distribution of radiating particles with corresponding upward-sloped $\epsilon F_{\rm IC}(\epsilon)$ emission spectrum peaking well above pair-production threshold (Fig. 9). Most of the initially emitted energy is, thus, veiled by pair production, and the peak of the apparent/observed spectrum lies instead just before the absorption-induced cut-off: that is, at energies $\sim\gamma_{\rm KN}m_ec^2$, as seen in the IC(KN)+PP run. Subsequently, in between reconnection-powered acceleration events, the radiative physics takes over in determining the mean photon energy. As seen during these periods in the Thomson regime, the IC spectrum softens because rapidly cooling particles radiating at the spectral peak energy, $\sim\gamma_{\rm rad,T}^2\,\epsilon_{\rm ph}$ (Fig. 9), cannot be replenished by electromagnetic energization. The key difference in the IC(KN)+PP case is that particles emitting at the observed peak energy, $\sim\gamma_{\rm KN}m_ec^2$, can still be partially replenished by IC cooling and pair production, which actively reprocess radiation originally emitted at higher, absorbed photon energies down to the observed band. This stabilizes the observed spectrum, even in the falling phase of the light curve when electromagnetic energization is completely absent. Hence, the photon energy $0.3\gamma_{\rm KN}m_ec^2$ owes its luminosity-independent stability to the fact that the coupling of reconnection-powered NTPA to Klein–Nishina radiative physics results in the same natural photon energy scale as that produced by the radiative physics alone.

We note that radiative reprocessing of initially above-threshold photons also leads, after reconnection has concluded, to the self-similar t^{-6} power-law brightness decay shown in Fig. 12. Although this is much slower than in the case of Thomson radiative cooling, it is still probably too abrupt for gamma-ray instruments to resolve. Thus, what we would like to stress as the main difference between gamma-ray flares powered by Thomson-radiative and Klein–Nishina reconnection is that the latter are characterized by *a constant mean observed photon energy, irrespective of brightness*.

We conduct post-processing experiments in Fig. 13 that suggest that the presented properties of the IC(KN)+PP light curve and spectrum in the absence of particle acceleration are universal. In these experiments, we evolve different initial power-law distributions of particles – $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$ for $p = 1$, 2, and 3, and $\gamma_{\rm cool} < \gamma < \gamma_{\rm max}$ (dashed lines in Fig. 13, left panel) – solely under the influence of Klein–Nishina emission and pair production (as in the post-processing phase of Fig. 12 except that, there, the initial particle distribution is taken from time t_0 of our PIC simulation). We find that, irrespective of the initial power-law slope, the particle distribution always relaxes, in the Thomson regime, $\gamma \lesssim \gamma_{\rm KN}$, to a γ^{-2} scaling (solid lines in Fig. 13, left panel).[6] This corresponds,

[6]The slope, γ^{-2}, can be calculated by considering the trickle of particles from $\gamma > \gamma_{\rm KN}$ as monochromatic particle injection at $\gamma = \gamma_{\rm KN}$ (Mehlhaff et al. 2021).

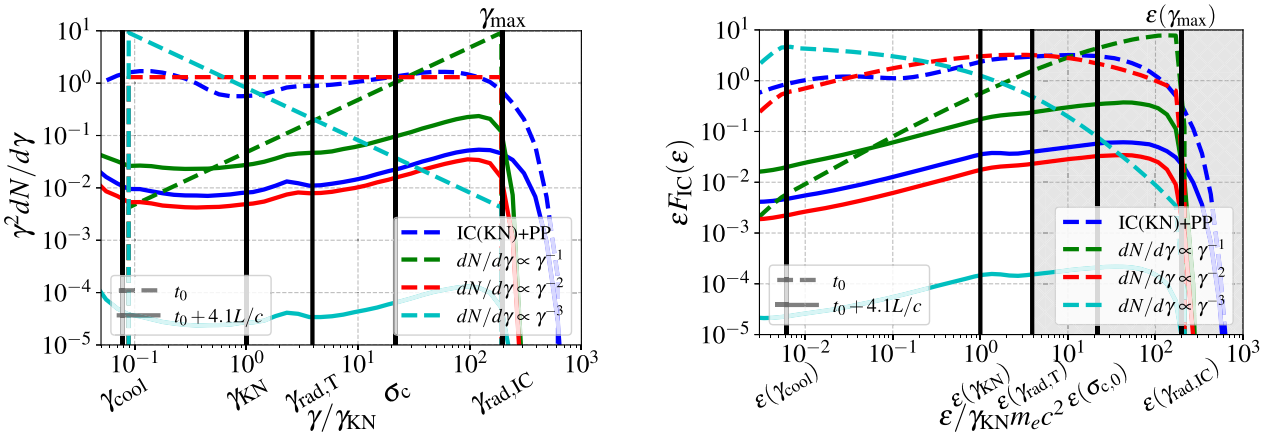

Figure 13. Left: numerical experiment wherein different initial power-law particle distributions, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$ for $p = 1$, 2, and 3, are passively cooled under IC emission and pair production. All parameters are identical to the IC(KN)+PP simulation. Integration is done from time t_0 (dashed curves) to $t_0 + 4.1L/c$ (solid curves). For reference, we present, in blue, the passive cooling of the particle distribution yielded by the IC(KN)+PP simulation (as in Fig. 12). Right: IC emission spectra corresponding to the particle distributions at left. A grey region indicates unobserved (above pair-production threshold) photon energies. After an initial-condition-dependent relaxation period, all distributions converge to an initial-condition-independent universal shape. At observable energies – that is, in the Thomson regime, $\epsilon \lesssim \gamma_{\rm KN} m_e c^2$ ($\gamma \lesssim \gamma_{\rm KN}$) – this universal shape features an approximate γ^{-2} power law, corresponding to $\epsilon F_{\rm IC}(\epsilon) \propto \epsilon^{0.5}$. (The apparent upturn in the particle distributions at low energies is an artefact of short integration times – it results from a bulk population of cooled particles that would gradually become still colder with time, extending the γ^{-2} scaling to even lower energies.).

for the Thomson part, $\epsilon \lesssim \gamma_{\rm KN} m_e c^2$, of the emission spectrum, to $\epsilon F_{\rm IC}(\epsilon) \propto \epsilon^{-(2-3)/2} = \epsilon^{1/2}$ (see Section 4.4) – a rising spectrum that continues almost up to the pair-production threshold energy $\epsilon = 4\gamma_{\rm KN} m_e c^2$ (solid lines in Fig. 13, right panel). These universal shapes, once reached, are maintained by the particle distribution and emission spectrum as they fall off, resulting in a self-similar t^{-6} luminosity decay law and constant mean photon energy, $\langle\epsilon\rangle \sim 0.3\gamma_{\rm KN} m_e c^2$: the same as Fig. 12.

The above exercise enables us to reason about the observable signatures of Klein–Nishina reconnection in regimes, not probed by our simulations, where reconnection-powered NTPA is known to yield a steeper particle energy distribution. This occurs, for example, in the presence of a strong guide field (Werner & Uzdensky 2017) or in the transrelativistic regime of electron–proton plasmas (Werner et al. 2018). In such cases, we speculate that one would observe an initially steep emission spectrum corresponding to intrinsic reconnection-powered NTPA, followed, in the passive cooling phase, by a transition – in fact, a hardening! – to the identified universal shape as the flux decays (similar to the $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-3}$ initial condition in Fig. 13). This might be difficult to observe, however, as the spectrum may dim too much before relaxing to the expected shape, starving gamma-ray instruments of a sufficient number of photons to reconstruct it (cf. the large gap between the initial and final $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-3}$ curves in Fig. 13).

We remark that the universal shape of the passively evolved (i.e. only via IC losses and pair production but without particle acceleration) particle energy distribution – and the consequent stability of the emission spectrum when energization subsides – depends only on Klein–Nishina effects that are independent of the number of spatial dimensions. This result should therefore generalize to 3D reconnection. However, given the differing acceleration physics in 3D (e.g., as discussed in Section 4.3, particles escaping from the reconnection downstream and subsequently being re-accelerated, Zhang et al. 2021; Chernoglazov et al. 2023; Zhang et al. 2023), the stability of the spectrum on the rising part of the light curve may change. Ultimately, the 3D behaviour needs to be verified by full-fledged 3D simulations.

In this section, we have seen how the different regimes of radiative cooling treated by this study – particularly their influence on particle acceleration and the resulting IC emission spectra – give rise to highly distinct temporal radiative signatures. At the most coarse-grained level, emission from radiative reconnection tracks the electromagnetic dissipation in real time, whereas emission from non-radiative reconnection traces only energization that has occurred in the past. Focusing on more specific observable differences in the context of reconnection-powered flares, the light curve from Thomson radiative reconnection is highly correlated with the observed average photon energy (provided one observes near the spectral peak at $\sim \gamma_{\rm rad,T}^2 \epsilon_{\rm ph}$) and features an abrupt decay phase where both drop simultaneously. Klein–Nishina radiative reconnection with pair production, however, yields an exactly opposite trend, with no correlation between the average photon energy, $\sim 0.3\gamma_{\rm KN} m_e c^2$, and the overall brightness, including in the (slower than in the Thomson regime, but still relatively fast) decay phase. These findings can be directly compared with, and tested by, observations of gamma-ray flares from suitable astrophysical systems (see Section 6). Finally, though this section is predicated on simulations with periodic boundaries and, hence, well-defined beginning and end phases of reconnection, the main results are likely to be more general. In particular, the same arguments used to interpret the relative spectral stability of Klein–Nishina reconnection compared to the Thomson case still hold in steady-state reconnection. Whether such effects would be observable in a steady state, however, depends on: (1) whether sufficiently large statistical fluctuations about the time average would occur to yield detectable variability in the observed radiation; and, (2) whether such variability would be distinguishable from that owing to quasi-static changes to the large-scale reconnection layer, as may be caused by macroscopic dynamics of the astrophysical system (cf. Section 6.4).

4.6 Newborn pair energy budget and particle count

In the preceding parts of Section 4, we explored consequences of Klein–Nishina and pair-production physics on magnetic reconnec-

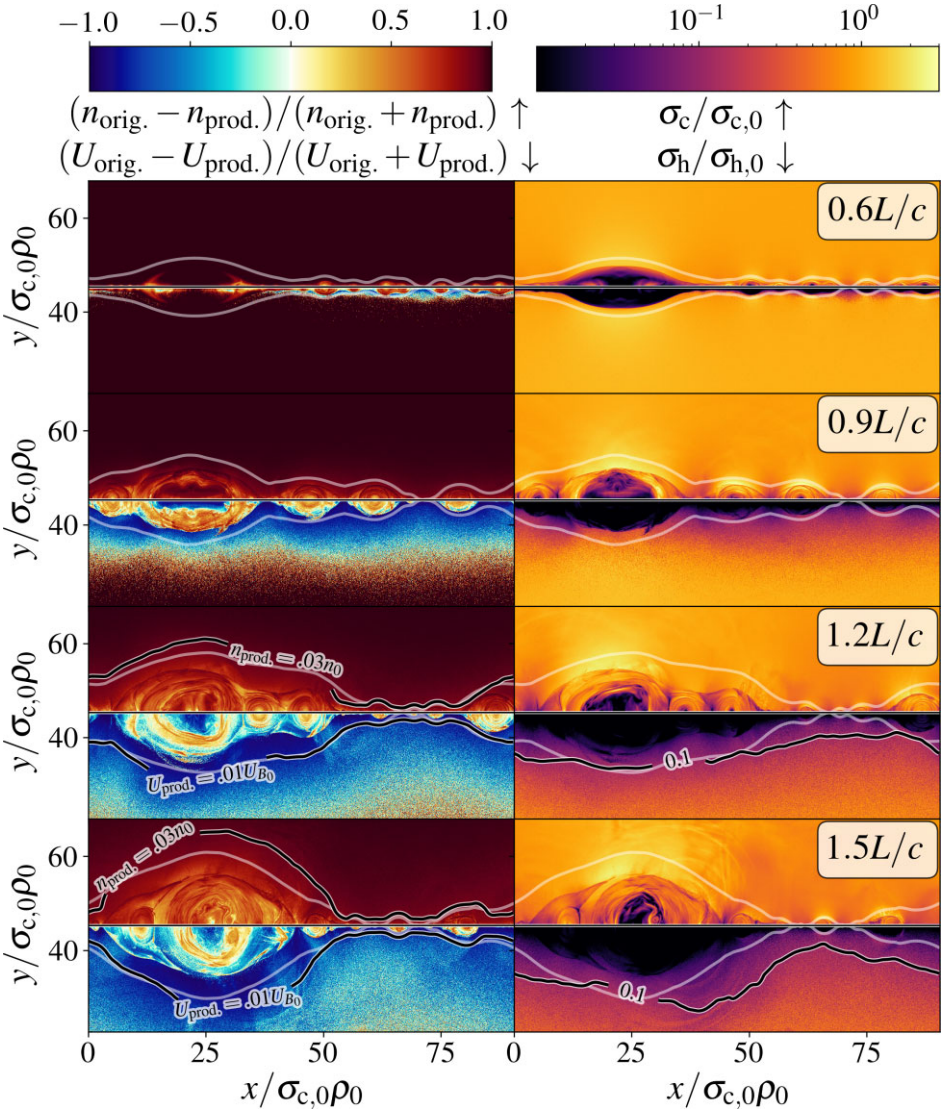

Figure 14. Left-hand column: snapshots of the relative contributions of original (orig.) and produced (prod.) pairs to (on the top half of each panel) the local particle number density, $(n_{\rm orig.}-n_{\rm prod.})/(n_{\rm orig.}+n_{\rm prod.})$, and (on the bottom half of each panel) the local particle energy density, $(U_{\rm orig.}-U_{\rm prod.})/(U_{\rm orig.}+U_{\rm prod.})$. Right-hand column: snapshots of the local cold, $\sigma_{\rm c}/\sigma_{\rm c,0}$ (top halves of panels), and hot, $\sigma_{\rm h}/\sigma_{\rm h,0}$ (bottom halves of panels), magnetizations. Contours display values of key quantities along, or just upstream of, the reconnection separatrix. For the $U_{\rm prod.}$ contour, $U_{B_0}\equiv B_0^2/8\pi$. In agreement with Mehlhaff et al. (2021), newborn pairs are everywhere subdominant in terms of their number density, but compete with the original particles for energetic dominance. Moreover, the newborn particles create an energetically dense coat around the reconnection layer where they load the local hot magnetization but not the cold magnetization.

tion, using as control cases the Thomson-radiative and non-radiative regimes. We presented first what is similar to the latter two cases – for example, the overall spatial dynamics and the reconnection rate – and then discussed the main distinctions, culminating with the very different observable signatures of the various radiative regimes. We now go one step farther, leaving behind our control cases in order to address issues that only exist in the context of Klein–Nishina reconnection with pair production. In particular, we comment on the newborn pairs' contribution to the reconnection system's energy and particle number budgets, which is presented graphically in Fig. 14.

The left-hand column of Fig. 14 illustrates the local differences in number densities, $(n_{\rm orig.}-n_{\rm prod.})/(n_{\rm orig.}+n_{\rm prod.})$, and energy densities, $(U_{\rm orig.}-U_{\rm prod.})/(U_{\rm orig.}+U_{\rm prod.})$, between originally present (orig.) and produced (prod.) particles. In the number density panels, the newborn particles remain subdominant across time and space. However, once reconnection ignites above-threshold gamma-ray emission and pair production, the produced particles build up an energetically dense coat around the reconnection layer. Inside this coat, the newborn pairs completely dominate the energy density of the reconnection upstream region; in the reconnected flux region, while not as clearly dominant, they still vie for control of the energy budget. This demonstrates the main peculiarity of Klein–Nishina reconnection – previously anticipated by Mehlhaff et al. (2021) – that, for a broad range of parameters (made more precise below), the produced particles are hot but tenuous, competing with the original particles for energetic dominance of the system despite being much fewer in number. A parallel view of these effects is presented in the right-hand column of Fig. 14, where we plot the local cold magnetization $\sigma_{\rm c}\equiv(B_x^2+B_y^2)/[4\pi(n_{\rm orig.}+n_{\rm prod.})m_ec^2]$ and the hot

magnetization $\sigma_{\rm h} \equiv (B_x^2 + B_y^2)/4\pi w$ [note: the enthalpy density, for our ultrarelativistic particle distributions, is $w = 4(U_{\rm orig.} + U_{\rm prod.})/3$; see equation (32) and surrounding discussion]. Because the newborn pairs load the upstream plasma energy density without appreciably changing the number density, they reduce $\sigma_{\rm h}$ from its native/initial value, $\sigma_{\rm h,0}$, while leaving $\sigma_{\rm c} \simeq \sigma_{\rm c,0}$ essentially untouched.

Even though the newborn pairs are everywhere less numerous than those originally present, they discernibly contribute to the plasma density [e.g. by changing the hue of the $(n_{\rm orig.} - n_{\rm prod.})/(n_{\rm orig.} + n_{\rm prod.})$ spatial maps] close to and inside of the reconnection separatrix. Mehlhaff et al. (2021) forecasted that the newborn pairs should begin to contribute non-negligibly to the plasma number density passing across the separatrix when $\sigma_{\rm c,0}$ exceeds $\gamma_{\rm KN}$ by more than a factor of 10 or so. This is consistent with our findings on the Klein–Nishina reconnection pair yield presented in the next section, which suggest that the IC(KN)+PP run ($\sigma_{\rm c,0} \simeq 20\gamma_{\rm KN}$) is indeed beginning to border on a regime where the *in situ* produced pairs contribute more significantly to the particle count. For the rest of this section, however, we focus on the present regime where the newborn particles are energetically dense but numerically few.

As discussed by Mehlhaff et al. (2021), the energy density injected into the upstream region via the hot newborn pairs leads to a *pair-loaded magnetization*, $\sigma_{\rm h}^*$, to which the system tends to regulate itself (provided the initial magnetization, $\sigma_{\rm h,0}$, exceeds $\sigma_{\rm h}^*$). However, if the pair loading is efficient enough, the system may never actually attain a steady state with $\sigma_{\rm h} = \sigma_{\rm h}^*$. It will instead flood the upstream energy density via pair production, overshooting to a lower magnetization, $\sigma_{\rm h} < \sigma_{\rm h}^*$, and quenching subsequent NTPA until the upstream plasma is vacated and a high $\sigma_{\rm h} > \sigma_{\rm h}^*$ is restored. This would restart the process, leading to a limit cycle: the system would indefinitely ricochet between a high and a low magnetization on either side of $\sigma_{\rm h}^*$ throughout the duration of reconnection.

Mehlhaff et al. (2021) quantified the dependence of the long-term fate of the system – whether it smoothly regulates to, or violently oscillates about, $\sigma_{\rm h}^*$ – on the efficiency of upstream pair loading. We measure this efficiency from our simulations in Appendix B, finding that it is too low, based on the analysis of Mehlhaff et al. (2021), to trigger $\sigma_{\rm h}$-mediated limit cycles. We therefore conclude that *limit cycles mediated by pair-loading of σ_h are unlikely in astrophysical Klein–Nishina reconnection*.

In this section, we have shown that our simulations probe a Klein–Nishina reconnection regime where, as previously predicted by Mehlhaff et al. (2021), the newborn upstream pairs are hot but tenuous, loading the upstream plasma pressure, energy density, and hot magnetization, but not the upstream plasma density or cold magnetization. The simulations further provide measurements (Appendix B) that disfavour $\sigma_{\rm h}$-mediated limit cycles. However, if a regime existed featuring a large multiplicity of newborn pairs, the possibility of $\sigma_{\rm c}$-mediated limit cycles would remain an open question. It is the overall pair yield of Klein–Nishina reconnection, including the potential existence of such a copious pair-production regime, to which we now turn.

5 PAIR YIELD

We have already seen how the pairs produced in Klein–Nishina reconnection contribute unique aspects to its observable signatures and self-consistent internal dynamics. In addition to these intrinsic features, another important implication of Klein–Nishina reconnection is its interaction with its environment. Here, pair production opens up a coupling channel that is unique to QED reconnection: the possibility to change the ambient plasma composition (e.g. the positron-to-proton ratio). Thus, in this section, we characterize the total pair yield from a Klein–Nishina reconnection event.

To map the dependence of the pair yield on reconnection parameters, we introduce two auxiliary simulation campaigns. Each one explores the pair yield's dependence on one principal variable. The first campaign explores the main new quantity introduced by Klein–Nishina physics, $\gamma_{\rm KN}$. The second campaign is a system-size (L) scan. It doubles as an opportunity to diagnose how well our results may generalize to astrophysical Klein–Nishina reconnection, where the layer lengths, L, are expected to be much larger compared to the plasma microscales (e.g. $\sigma_{\rm c,0}\,\rho_0$) than is possible to simulate. The full sets of parameters used for the respective $\gamma_{\rm KN}$- and L-scans are summarized in Tables 2 and 3.

Besides examining the impact of their respective variables, these additional sets of simulations are separately calibrated to different fiducial parameters. The $\gamma_{\rm KN}$-scan has a larger normalized system size, $L = 210\sigma_{\rm c,0}\rho_0$, and lower magnetizations, $\sigma_{\rm c,0} = 10^4$ and $\sigma_{\rm h,0} = 10^2$, than our base run, IC(KN)+PP (for which $L \simeq 91\sigma_{\rm c,0}\rho_0$, $\sigma_{\rm c,0} = 1.2 \times 10^5$, and $\sigma_{\rm h,0} = 1250$; Table 1). The L-scan has still

Table 2. Parameters for the six-simulation scan over $\gamma_{\rm KN}$ discussed in section 5. Values that are the same as those of Table 1 are omitted. For computational expediency, we employ a larger time-step in these runs (cf. Tables 1 and 3). The energy error is thus slightly worse, but is still only above 1 per cent for the $\gamma_{\rm KN} = 1500$, 1900, and 3400 runs, where it is, respectively, 1.2 per cent, 1.3 per cent, and 1.7 per cent.

Symbol	Value	(= equivalent)
$\sigma_{\rm c,0}$	10^4	
$\sigma_{\rm h,0}$	10^2	
θ_0	$2.5 \times 10^{-4}\sigma_{\rm c,0}$	$=25$
L	$210\sigma_{\rm c,0}\rho_0$	
$\gamma_{\rm max}$	$21\sigma_{\rm c,0}$	$=2.1 \times 10^5$
B_g	$0.1B_0$	
$\Delta x,\ \Delta y$	$\sigma_{\rm c,0}\rho_0/24$	
Δt	$0.99\Delta x/\sqrt{2}c$	
N	5120	
$\gamma_{\rm cool}$	$2.3 \times 10^{-2}\sigma_{\rm c,0}$	$=230$
$\gamma_{\rm rad,\,T}$	$0.70\sigma_{\rm c,0}$	$=7.0 \times 10^3$
$\gamma_{\rm KN}$	$(0.091, 0.12, 0.15)\sigma_{\rm c,0}$	$=$ (910, 1200, 1500)
...	(0.19, 0.27, 0.34)$\sigma_{\rm c,0}$	=(1900, 2700, 3400)
$\tau_{\gamma\gamma}$	(2.4, 3.1, 4.0, 5.2, 7.2, 9.1)	

Table 3. Parameters for the three-simulation scan over L. Values that are the same as those of Table 1 are omitted. The value of $\gamma_{\rm KN}$ is adjusted slightly between runs to keep $\gamma_{\rm rad,\,IC} = \gamma_{\rm max}$.

Symbol	Value	(= equivalent)
$\sigma_{\rm c,0}$	400	
$\sigma_{\rm h,0}$	400*	
θ_0	0.1	
L	$(34, 67, 100)\sigma_{\rm c,0}\rho_0$	
$\gamma_{\rm max}$	$(3.4, 6.7, 10.)\sigma_{\rm c,0}$	=(1300, 2700, 4000)
B_g	$0.1B_0$	
$\Delta x,\ \Delta y$	$\sigma_{\rm c,0}\rho_0/76$	
N	(2560, 5120, 7680)	
$\gamma_{\rm cool}$	$(12, 5.9, 4.0) \times 10^{-3}\sigma_{\rm c,0}$	=(4.7, 2.4, 1.6)
$\gamma_{\rm rad,\,T}$	$0.2\sigma_{\rm c,0}$	=80
$\gamma_{\rm KN}$	$(0.065, 0.055, 0.051)\sigma_{\rm c,0}$	=(26, 22, 20.)
$\tau_{\gamma\gamma}$	(3.3, 5.6, 7.8)	

Note. * Because the initial plasma temperature, $\theta_0 = 0.1$, is non-relativistically cold, $\sigma_{\rm h,0} = \sigma_{\rm c,0}$.

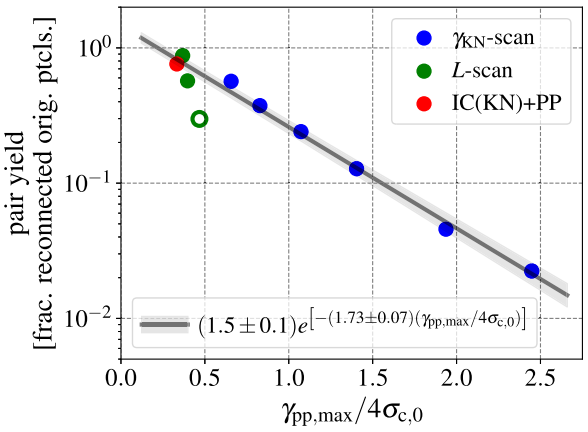

Figure 15. Pair yield for the simulations of the γ_{KN} parameter scan (blue; Table 2), the L parameter scan (green; Table 3), and the IC(KN)+PP run (red; Table 1). The smallest-L simulation (open circle) is omitted when fitting for the exponential scaling law (solid line with error envelope indicating reported 1σ uncertainty on fit parameters). Despite the large differences in their fiducial and scanned parameters, all simulations' pair yields (with the exception of the smallest one in the L-scan; see the text) fall onto a scaling law with a single control parameter. This parameter is the ratio of two Lorentz factors: that, $\gamma_{pp,\,max} \simeq 30\gamma_{KN}$, of a particle that typically scatters photons up to the peak energy of the pair-production cross-section; and the characteristic maximum Lorentz factor, $\gamma_X \equiv 4\sigma_{c,0}$, attained near reconnection X-points. This suggests that the pair yield is controlled mainly by the efficiency of particle acceleration up to and beyond the optimal pair-producing particle energy.

different initial magnetizations, $\sigma_{c,0} = \sigma_{h,0} = 400$, and features a non-relativistic initial upstream plasma, $\theta_0 = 0.1$. Spreading out, in this way, our auxiliary campaigns around the Klein–Nishina reconnection parameter space helps us identify a reduced set of control parameters (in fact, one single parameter; Fig. 15) that decides the pair yield. These control parameters, in turn, shed light on the main physical mechanisms responsible for the pair-production efficiency while also providing a potential method for estimating this efficiency in astrophysical systems.

We define the pair yield in our simulations as the ratio of the total number, $N_{prod.}$, of leptons (electrons and positrons) produced on the fly to the cumulative count, $N_{rec.}$, of originally present leptons processed by reconnection (i.e. swept across the separatrix). Generally, about 60 per cent of the initial upstream particles cross the separatrix before reconnection saturates, the same as the percentage of the initial magnetic flux that is reconnected (see Section 4.2 and Fig. 5). Though the processing of upstream magnetic field and original particles essentially finishes by the end of our simulations, pair production does not; there remains a prominent population of high-energy particles that have yet to cool down by emitting pair-producing gamma-rays. In Section 4.5, we post-processed the passive cooling associated with these particles to diagnose the decaying phase of the reconnection-powered light curve. Here, we use the same technique to continue evolving the total newborn pair count past the end of each simulation. Once the count saturates (typically by $6L/c$ or so), we record it as $N_{prod.}$. As long as we begin the post-processing after the energy transfer from fields to particles is mostly complete, the final $N_{prod.}$ does not depend much (less than 10 per cent) on the exact moment in time when the post-processing starts.

Let us now comment on the generality of results obtained via the above pair yield measurement recipe. We note that the measurement procedure is easily formulated and performed in the context of our periodic simulation box. The fact that the processing of magnetic flux, advection of upstream particles across the separatrix, and particle energization all eventually slow down and cease enables a straightforward calculation of the pairs created per reconnection-processed original particle. This does not necessarily limit the applicability of the pair yield of this section to flaring sources with discrete reconnection episodes, however. On the contrary, if the pair yield is truly local, scaling linearly – for a fixed set of parameters (those outlined in Section 3) – with the number of processed upstream particles, then the results would also hold in the context of steady-state reconnection.

As a potential caveat, one should note that, like our periodic simulation boundaries, our pair yield measurement technique ignores the potential for particles and photons to escape the ambient radiation field before pair production is complete. This issue is less important when the pair-production optical depth, $\tau_{\gamma\gamma}$, is large, corresponding to shorter mean-free paths of photons and more rapid particle cooling times (Fig. 2 and Section 3.2.1). When $\tau_{\gamma\gamma}$ becomes small, the pair yield calculated in this way still has meaning as long as the extent of the ambient radiation field is much larger than the size of the reconnection system (and provided, once particles exit the reconnection system, they are no longer significantly energized).

Having discussed the general applicability and limitations of our pair yield measurements, we now present the pair yield calculated for our auxiliary γ_{KN} and L simulation campaigns, as well as for the single IC(KN)+PP run discussed earlier (Section 4), in Fig. 15. Remarkably, when plotted as a function of just the single control parameter, $\gamma_{pp,\,max}/4\sigma_{c,0}$, the pair yields from all simulations – despite the very different fiducial and scanned parameters – collapse onto the same exponential scaling law,

$$N_{prod.}/N_{rec.} = 1.5\exp\left(-1.7\gamma_{pp,max}/4\sigma_{c,0}\right) . \tag{53}$$

Here, $\gamma_{pp,\,max} \equiv 3.6 \times 8 \times \gamma_{KN} \simeq 30\gamma_{KN}$ is the characteristic Lorentz factor of particles that scatter background photons to energies at peak pair-production cross-section, $\epsilon_{scat} \sim \gamma_{pp,\,max}/2 \sim 3.6(m_e c^2)^2/\epsilon_{ph}$. Meanwhile, $4\sigma_{c,0}$ is the characteristic maximum energy, $\gamma_X \equiv 4\sigma_{c,0}$, that particles acquire near reconnection X-points. Equation (53) suggests that what controls the final pair yield of reconnection is how broad a distribution of high-energy particles can be energized near reconnection X-points to radiate photons close to or above the peak pair-production cross-section: that is, by how much γ_X exceeds $\gamma_{pp,\,max}$.

Mehlhaff et al. (2021) predicted that the ratio of the newborn-to-original upstream pair densities flowing across the reconnection separatrix should be proportional to $\sigma_{c,0}/\gamma_{KN}$ (times a non-trivial function depending on NTPA in the reconnection layer). Although this number density ratio is not the same as the global ratio of newborn-to-reconnection-processed particles measured here – the latter also includes the non-negligible number of pairs born on the exhaust side of the reconnection separatrix – both results share the same main controlling parameter, $\sigma_{c,0}/\gamma_{KN} \propto \gamma_X/\gamma_{pp,\,max}$.

Now let us discuss where the scaling (53) may break down. First, all of the simulations in our campaigns have an effective radiative cut-off Lorentz factor, $\gamma_{rad,\,IC} \geq 4\sigma_{c,0}$. This means that we need to be vigilant as γ_{KN} increases (coinciding with larger $\gamma_{pp,\,max}/4\sigma_{c,0}$), because $\gamma_{rad,\,IC}$ may then fall below $4\sigma_{c,0}$. In that case, our present interpretation of equation (53) leads us to expect another cut-off in the pair yield, for example, of the form, $\exp(-\gamma_{pp,max}/\gamma_{rad,IC})$, to kick in, signalling that X-point particle acceleration has become limited by $\gamma_{rad,\,IC}$ instead of by $\sigma_{c,0}$. Second, in the opposite limit, where $\gamma_{pp,\,max}/4\sigma_{c,0}$ becomes small, we are likely to transition to a regime in which extremely efficient particle acceleration (giving a particle

distribution power-law index approaching the non-radiative limit, $p \simeq 1$) causes the pair yield to continue to grow with $4\sigma_{c,0}/\gamma_{pp,max}$. Though this is not what a naive extrapolation of (53) produces, it is more coherent with our present interpretation of that formula, in which more particles being accelerated into the optimal pair-producing energy range enhances the pair yield.

Let us examine, for the sake of argument, what might happen if pushing $\gamma_{pp,max}/4\sigma_{c,0}$ to smaller values than studied here indeed led to a copious pair yield. Then Klein–Nishina reconnection would begin to move out of the regime of our simulations (discussed in Section 4.6 and Appendix B) where the produced particles are hot but tenuous. Instead, the newborn particles would become hot and abundant, potentially even loading the upstream cold magnetization to a reduced value, $\sigma_c^* < \sigma_{c,0}$, whereas in our simulations they modify only the hot magnetization (Section 4.6). If a steady state were to exist in such a regime, the cold magnetization would need to self-adjust until the pair yield as defined in equation (53) attains unity (times some efficiency factor to account for the fact that not all of the produced particles would be born into the upstream region; cf. Appendix B). According to Fig. 15, this occurs when $\gamma_{pp,max}/4\sigma_c^* \simeq 0.25$, or, equivalently, $\sigma_c^* \simeq \gamma_{pp,max} \simeq 30\gamma_{KN}$. Such a result would be highly attractive, for it would open up the possibility of measuring the lepton material composition (via σ_c^*) in terms of the seed photon energy, a much easier quantity to infer observationally.

Lastly, we note that the trend identified in Fig. 15 is broken at smaller system sizes. Indeed, we excluded the pair yield measurement from our $L = 34\sigma_{c,0}\,\rho_0$ simulation when fitting for equation (53). The break from the formula in smaller systems reflects findings of earlier simulation studies that collisionless reconnection transitions to an asymptotically large-system limit, corresponding to the multiple X-point, plasmoid-mediated regime, only once $L \gtrsim 40\sigma_{c,0}\rho_0$ (Werner et al. 2016). When respecting this limit, our simulations overlap the identified pair yield trend. However, we cannot rule out larger systems yielding even more efficient pair production than (53).

In this section, we have characterized the pair yield of Klein–Nishina reconnection in terms of a 1D exponential scaling law, equation (53) – despite the high-dimensional parameter space of this problem. This scaling law appears to be robust across an order of magnitude or more in γ_{KN}, $\sigma_{h,0}$, $\sigma_{c,0}$, and θ_0, while being respected across a factor of 3 or 4 in system size (the most that we can afford to probe in the large-system, $L > 40\sigma_{c,0}\rho_0$, regime given the stringent parameter constraints of the problem; cf. Section 3). Equation (53) may need to be modified in the transition region between Klein–Nishina ($\gamma_{KN} < \gamma_{rad,T}$) and Thomson ($\gamma_{KN} > \gamma_{rad,T}$) radiative reconnection, and it may also give way to a new regime of efficient (much greater than order-unity) pair yield when $\gamma_{pp,max}/4\sigma_{c,0}$ becomes much smaller than the values we test. Nevertheless, we are able to capture an order of magnitude in the control parameter, $\gamma_{pp,max}/4\sigma_{c,0}$, including near the point, $\gamma_{pp,max}/4\sigma_{c,0} \simeq 0.25$, where the pair yield reaches 1.

6 DISCUSSION

In this section, we discuss the relevance of our findings to gamma-ray observations of selected astrophysical systems: FSRQs, black hole accretion disc coronae, the M87* magnetosphere, and gamma-ray binaries. In each case, we argue why the operation of Klein–Nishina reconnection in these systems is expected on theoretical grounds. We further discuss consequences, for each system, of the results of Sections 4.5 and 5. For reference, we briefly recapitulate those results here in the context of potential links to observations.

The main finding of Section 4.5 is the marked departure of the observable signatures of radiative reconnection in the Klein–Nishina regime (with pair production) from the Thomson-cooled regime (without pair production). Namely, while Thomson-cooled reconnection features a tight correlation between the mean observed photon energy and the system's total luminosity (i.e. 'harder-when-brighter'), Klein–Nishina reconnection breaks this correlation, featuring a steady mean photon energy irrespective of the luminosity. This includes the decaying phase of a flaring event, wherein the average photon energy is preserved even as the luminosity drops. These findings serve therefore as an observational diagnostic. They can be directly compared to gamma-ray observations to build a case (or not) for the operation of Klein–Nishina reconnection in a given object. In contrast, the results of Section 5 function as an inference criterion, providing a method to estimate a quantity that is difficult to constrain from observations – the emitting region's material composition – using quantities that may be more readily measured or estimated (specifically, the magnetization, $\sigma_{c,0}$, and the Lorentz factor, $\gamma_{pp,max}$, of particles whose photons are at maximum pair-production cross-section with the radiation bath).

As an aid to navigating the astrophysical diversity treated by the following discussion, each system-specific section (6.1–6.4) is further subdivided into three parts: an overview of pertinent background information, the principal discussion of connections to this work, and a brief object-specific synopsis. Following our system-by-system presentation, we summarize our broad conclusions across all systems in Section 6.5 and Table 4. Readers interested in a particular object may skip directly to the corresponding subsection; those seeking a more general overview may wish to skip first to the end summary material (Section 6.5).

6.1 Flat-spectrum radio quasars and other blazars

6.1.1 Background information

Blazars are AGNs that launch bipolar relativistic jets, one of which (hereafter, the singular 'jet') travels toward the Earth. The jet's relativistic motion Doppler boosts its emission, leading to dramatic observable consequences. For example, blazars dominate the discrete sources on the extragalactic gamma-ray sky (e.g. Wakely & Horan 2008; Abdollahi et al. 2020) and, in the optical band, they routinely outshine the cumulative starlight of their host galaxies (Olguín-Iglesias et al. 2016). Blazar jet emission is also exceptionally broad, extending from radio frequencies up to gamma-rays in a characteristic non-thermal double-humped spectrum (Fossati et al. 1998; Ghisellini 2011; Ghisellini et al. 2017). The lower energy spectral hump originates from synchrotron radiation by relativistic electrons and positrons (henceforth 'leptons') spiraling around magnetic field lines in the jet. The higher-energy peak is frequently attributed to IC radiation also by relativistic jet leptons (e.g. Prandini & Ghisellini 2022).

Blazars are phenomenologically decomposed into two main subdivisions: FSRQs and BL Lacs.[7] FSRQs exhibit lower energy spectra, with the synchrotron component peaking at infrared energies and the higher-energy IC component peaking in the MeV-to-GeV gamma-rays. In contrast, the maxima of the synchrotron and IC spectra

[7]For our purposes, we fold the extreme/ultrahigh-frequency-peaked BL Lacs (EHBLs/UHBLs) high-frequency-peaked BL Lacs (HBLs), intermediate-frequency-peaked BL Lacs (IBLs), and low-frequency-peaked BL Lacs (LBLs) into the BL Lac class.

Table 4. A graphic summary of Section 6.

(1) Object class	(2) KN hierarchy realized?	(3) Observational connection	(4) Limiting factor	(5) Explanation
FSRQs				
IC(BLR) scenario	Yes	Strong	–	*Fermi*-LAT observations are broadly consistent with anticipated temporal-spectral signatures of KN reconnection (Section 4.5; Figs 16–18)
IC(HDR) scenario	Yes	Limited	Instrumental	The upcoming CTA will provide enhanced temporal resolution at the relevant TeV energies, making possible more explicit comparisons with this work
Black hole accretion disc coronae	Yes	Limited	Sources	Sources are not bright enough in the relevant 10 + MeV range to probe temporal variability. However, the spectral cut-off in Cyg X-1 is potentially consistent with a KN reconnection model
M87* magnetosphere	Yes	–	Modelling	Need to account for synchrotron cooling in order to make relevant predictions
Gamma-ray binaries	Yes	Limited	Modelling	Orbital modulations of light curves at the relevant GeV energies necessitate global modelling

Notes. Column (1) indicates the object class (roughly one per Sections 6.1–6.4). Column (2) indicates whether basic theoretical estimates suggest that the Klein–Nishina reconnection scale hierarchy (50), $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, is achieved in the given system. Column (3) indicates our judgment of how firmly the results of this study – in particular the observable signatures of Klein–Nishina reconnection discussed in Section 4.5 – can be connected to presently available observational data. If strong observational connections cannot be made, column (4) indicates what, in our view, is the primary reason for this. Finally, column (5) gives a short explanation of the judgments in columns (3) and (4).

in BL Lacs typically lie in the UV/X-ray and GeV-to-TeV bands, respectively. Despite their lower photon energies, FSRQs are more luminous and exhibit much larger ratios of IC-to-synchrotron power. Finally, where the norm for BL Lacs is featureless non-thermal spectra, FSRQs usually exhibit prominent broad emission lines or quasi-thermal radiation at lower energies. These are usually attributed to emission by the underlying AGN accretion disc and to reprocessing of the accretion disc light by circumnuclear material. (Illustrative references pertaining to this entire paragraph include: Fossati et al. 1998; Ghisellini 2011; Madejski & Sikora 2016; Ghisellini et al. 2017; Blandford, Meier & Readhead 2019; Prandini & Ghisellini 2022.)

The circumnuclear regions that are observed at lower energies in FSRQs can provide intense sources of seed photons for IC emission in the jet (Begelman & Sikora 1987; Melia & Königl 1989; Sikora, Begelman & Rees 1994). Particularly bright are the broad emission line region (BLR) and the hot dust region (HDR). Of these two, the BLR is smaller, occupying an inner zone (up to roughly 0.1 pc from the nucleus) where irradiation from the accretion disc ionizes the ambient gas, and subsequent recombination emits line emission, broadened by rapid orbital motion, of characteristic UV energy

$$\epsilon_{\rm BLR} = 10\,{\rm eV}\,, \tag{54}$$

onto the jet (Tavecchio & Ghisellini 2008; Sikora et al. 2009; Nalewajko et al. 2012). The outer circumnuclear reprocessing region (up to roughly 4 pc from the nucleus) is the HDR, which comprises dust radiatively heated by the accretion disc light up to a temperature of about $T_{\rm HDR} \sim 1200\,{\rm K}$. The hot dust shines a quasi-thermal spectrum onto the jet of characteristic energy (Nenkova et al. 2008a, b; Sikora et al. 2009; Nalewajko et al. 2012)

$$\epsilon_{\rm HDR} \sim 3k_{\rm B}T_{\rm HDR} = 0.3\,{\rm eV}\,. \tag{55}$$

The radiation fields from the BLR and the HDR furnish excellent conditions for comparing with our simulations. First, they are energetically dense at distances far from the central engine such that the magnetic field energy density is small compared to that of the seed photons, $B_0^2/8\pi \ll U_{\rm ph}$, a necessary condition for neglecting synchrotron losses, as we do in our simulations. Second, the resulting radiation field is expected to be homogeneous not just across the reconnection region, but also across the whole jet width. This creates a direct opportunity for applying our pair yield law found in Section 5, which ignores the possibility of above-threshold photons escaping the ambient radiation field before being absorbed to produce electron–positron pairs.

We conduct a detailed analysis of scenarios where reconnection powers high-energy IC emission in FSRQ jets in our previous analytic work (Mehlhaff et al. 2021). There, we estimate the Lorentz factor energy scales $\gamma_{\rm cool}$, $\gamma_{\rm KN}$, $\gamma_{\rm rad,\,T}$, and $\gamma_{\rm max}$ either for the case where the reconnection region lies within the more energetically dense BLR, and thus leptons scatter primarily BLR photons, or for the case where the reconnection region is outside the BLR but inside the HDR such that the BLR radiation field is diluted and the HDR supplies the dominant seed photons. Our estimates in both scenarios yield fiducial energy scales that are in the required order, $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$ [equation (50)], to realize Klein–Nishina reconnection.

In Mehlhaff et al. (2021), we also pointed out that the BLR and HDR are optically thick to gamma-rays above the pair-production threshold energies

$$\epsilon_{\rm th,BLR} = m_e^2c^4/\epsilon_{\rm BLR} = 30\,{\rm GeV} \tag{56}$$

and

$$\epsilon_{\rm th,HDR} = m_e^2c^4/\epsilon_{\rm HDR} = 0.9\,{\rm TeV}\,, \tag{57}$$

respectively. The corresponding characteristic energies radiated by a particle of energy $\gamma_{\rm KN}$ in each case are $\gamma_{\rm KN,BLR}m_ec^2/2 = \epsilon_{\rm th,BLR}/8 = 3\,{\rm GeV}$ and $\gamma_{\rm KN,HDR}m_ec^2/2 = \epsilon_{\rm th,HDR}/8 = 0.1\,{\rm TeV}$. This means that observations by the *Fermi* Large Area Telescope (LAT), which is sensitive roughly to energies in the $0.1 - 100\,{\rm GeV}$ range (Atwood et al. 2009), are able to probe emission by particles at $\gamma_{\rm KN,\,BLR}$ up through the BLR gamma-ray absorption cut-off. At the same time, Imaging Atmospheric Cherenkov Telescopes (IACTs), typically sensitive in the $0.1 - 10\,{\rm TeV}$ band (Cherenkov Telescope Array Consortium et al. 2019), stand best to capture the analogous physics for the IC(HDR) scenario. This is fortuitous because particles with energies near $\gamma_{\rm KN}$ are precisely those responsible for the characteristic spectral and timing signatures of Klein–Nishina reconnection uncovered in

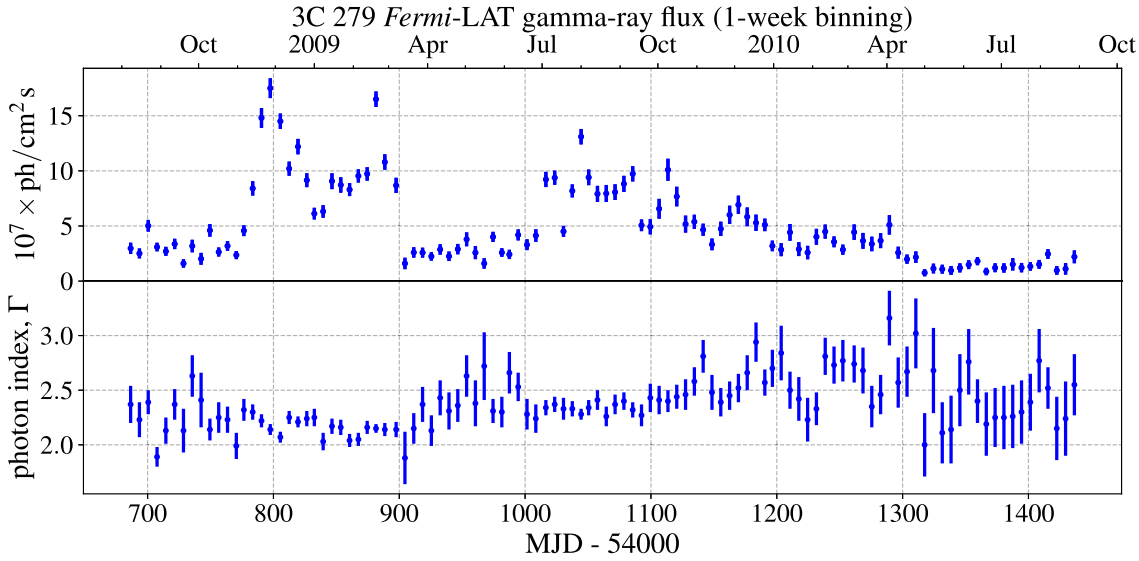

Figure 16. *Fermi*-LAT photon flux (top) and spectral index (bottom), both binned into one-week time intervals, detected from 3C 279 during the period presented by Hayashida et al. (2012). Error bars indicate symmetric Gaussian equivalent 1σ error. Top and bottom panels, correspond, respectively, to panels (c) and (f) of those authors' fig. 1. The data were retrieved from the *Fermi*-LAT Light Curve Repository (Abdollahi et al. 2023). The photon index, Γ, is defined such that the flux of photons between energies ϵ and $\epsilon + \mathrm{d}\epsilon$ is proportional to $\epsilon^{-\Gamma}$. This is connected to the $\epsilon F(\epsilon)$ representation, for example, of Fig. 9, in that $\epsilon F(\epsilon) \propto \epsilon^{-\Gamma+2}$.

Section 4.5. These particles radiate just below pair-production threshold, producing the highest energy observable photons, and they are actively replenished by radiative reprocessing from higher (above-threshold and, hence, invisible) photon energies, which stabilizes the observed spectral energy density. Thus, FSRQ flares in the GeV and TeV bands are ideally suited to probe the expected observational signatures of reconnection in the regime studied in this work.

6.1.2 *Connection to this work*

Let us examine a few instructive flares observed by the *Fermi*-LAT [which, again, probes the IC(BLR) scenario] from the FSRQ, 3C 279. Hayashida et al. (2012) report 3C 279 flaring periods over the first two years of *Fermi* operations. For reference, we reproduce the light curve and photon index time-series presented by those authors in their fig. 1, as well as the correlation between the total gamma-ray flux and spectral shape shown in their fig. 3, in our respective Figs 16 and 17 using data retrieved from the Fermi LAT Light Curve Repository (Abdollahi et al. 2023). Hayashida et al. (2012) note mild 'harder-when-brighter' behaviour over the entire observation period, but this is somewhat quenched during the brightest periods (Fig. 17), during which the *Fermi*-measured spectral index becomes remarkably flux-independent (e.g. between MJD 54 800 and 54 900 as well as between MJD 55 000 and 55 100 in Fig. 16). Roughly similar behaviour is often, but not uniformly, seen in later observations of the same object. In two even brighter outbursts from 3C 279 reported by Hayashida et al. (2015) and Ackermann et al. (2016), the gamma-ray flux reached high-enough levels to reconstruct spectra for individual orbits of the *Fermi* satellite. In the latter event, the flaring individual-orbit spectra reveal a photon energy index that remains between about 1.9 and 2.1 while the flux varies across about a factor of 3 (table 1 of Ackermann et al. 2016). The pre- and post-outburst phases of the event also feature rather stable photon indices (though this appears more statistically significant for the pre-outburst phase; see fig. 1 of Ackermann et al. 2016). As an example of when such spectral stability is not seen, the first flare analysed by Hayashida et al. (2015) exhibits an extreme hardening of the photon index that then softens on the trailing edge of the flare.

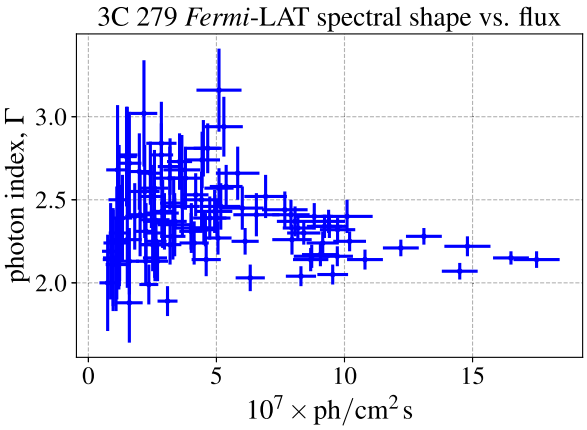

Figure 17. The data from Fig. 16 re-arranged into a scatter plot of gamma-ray spectral index versus photon flux. This figure roughly corresponds to fig. 3, panel (A) of Hayashida et al. (2012). The stabilization of the spectrum during high-flux periods, notably from MJD 54 800 to 54 900 and from MJD 55 000 to 55 100 in Fig. 16, is quite reminiscent of that found in Section 4.5 (cf. Fig. 12).

To provide a complete view of the correlation between spectral hardness and luminosity of 3C 279 in the GeV band, we also supply Fig. 18. This is similar to Fig. 17 except that the *Fermi*-LAT spectral index is plotted against the gamma-ray flux level for all 14+ yr of archived data. The full set of *Fermi* observations demonstrates broad consistency with, for example, the individual flaring period presented by Hayashida et al. (2012) and displayed in Figs 16 and 17: as the object brightens, the variation in photon index appears to decrease, suggesting a stabler spectrum during flaring periods.

At an even more general level, Meyer, Scargle & Blandford (2019) conducted a statistical analysis of the brightest flares from 6 of the most luminous FSRQs detected by *Fermi*. They find hints of 'harder-

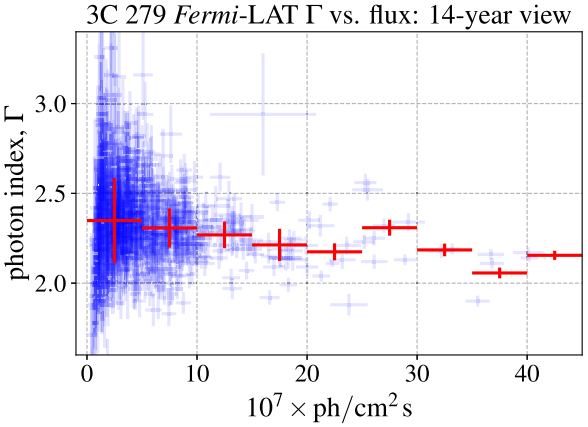

Figure 18. The same as Fig. 17, but plotting all (blue background crosses) of the one-week-binned LAT flux levels against corresponding spectral indices for over 14 yr of 3C 279 observations in the *Fermi*-LAT Light Curve Repository (Abdollahi et al. 2023). Additionally, we average the spectral indices (red foreground crosses) in flux bins of width 5×10^7 photons cm^{-2} s^{-1}. This bin width is indicated by horizontal red error bars. Vertical red error bars show 1σ uncertainty levels after bin averaging.

when-brighter' behaviour in some flares from some objects, but no statistical significance. At the same time, they report that higher flaring flux tends to coincide with reduced spectral variability. These remarks seem to be representative of other individual *Fermi*-detected FSRQ outbursts that are (quasi-)contemporaneous with TeV flares seen by one or more IACTs, including from PKS 1222+216 (Aleksić et al. 2011), PKS 1441+25 (Abeysekara et al. 2015), PKS 1510−089 (Ahnen et al. 2017), PKS 0736+017 (H. E. S. S. Collaboration 2020), and QSO B1420+326 (MAGIC Collaboration 2021). In these events, the GeV (sub-)flare often coincides with a relative stabilization of the GeV spectral index: a seeming decoupling between the flux and the spectral shape – including on the falling part of the flare (albeit the spectral index error bars are larger there). Nevertheless, one does see hints of 'harder-when-brighter' trends in some of the GeV (sub-)flares.

Thus, FSRQ flares in the *Fermi*-LAT sensitivity band exhibit broad, though perhaps imperfect, consistency with Klein–Nishina reconnection coupled to BLR Compton seed photons. The trends seem to persist even into the flare decay: if radiative losses reverted entirely to the Thomson regime, this would induce a strong (and unobserved) correlation between the gamma-ray spectral index and the flux in the decaying part of the flares. This suggests the role of Klein–Nishina effects, as unveiled in this work, in stabilizing the spectrum even as the light curve declines. For the cases where a harder-when-brighter trend appears more evident, this could be induced by weak synchrotron losses perturbing the dominant temporal-spectral behaviour induced by Klein–Nishina and pair-production physics.

One caveat to our association of GeV FSRQ flares with IC(BLR) Klein–Nishina reconnection is the occasional (quasi-)contemporaneous detection of TeV gamma-rays (a few examples of which are listed two paragraphs prior). For these cases, a simple one-zone emission model for both the GeV and TeV outburst places the emission region outside the BLR, which would otherwise absorb the TeV photons [equation (56)]. This caveat is made more severe by population studies of *Fermi*-detected FSRQs purely in the GeV band, which find no evidence for gamma-ray absorption by the BLR seed photons in the vast majority of objects (Costamante et al. 2018; Meyer et al. 2019). As noted by Costamante et al. (2018), however, even within a single zone framework, these constraints become less severe if one attributes the flaring emission to a structure (in our case, a reconnection layer) that comoves with the jet rather than a stationary feature (e.g. a standing shock). Then, over an observing period, $\Delta t_{\rm obs}$, of just 1 d (typically comparable to or shorter than variability time-scales identified by TeV FSRQ observations, with two notable exceptions: Aleksić et al. 2011 and H. E. S. S. Collaboration 2021), Doppler time compression of the light curve allows the emitting zone to travel a distance, $d \sim \Gamma_{\rm j}^2 c \Delta t_{\rm obs} \sim 10^{17}$ cm, from the central engine for a fiducial jet Lorentz factor, $\Gamma_{\rm j} = 10$. This is at the edge of the BLR, which only extends up to roughly 0.1 pc $\sim 10^{17}$ cm from the nucleus (Tavecchio & Ghisellini 2008; Sikora et al. 2009; Nalewajko et al. 2012; Mehlhaff et al. 2021), reducing the importance of absorption for the (potentially up to TeV) part of the emission produced at larger distances. In this view, attributing emission from longer flaring periods (e.g. as in Fig. 16) to IC(BLR)-coupled reconnection demands positing the ejection of multiple reconnecting structures from the central engine, as, for example, in striped-jet models (Giannios & Uzdensky 2019).

We next consider FSRQ flares at TeV energies observed by IACTs. Such events probe the the IC(HDR) scenario wherein reconnection couples to Klein–Nishina and pair-production physics mediated by the HDR seed photons. Suitable observations are much more difficult to obtain here. Not only do FSRQs shine intrinsically weakly in the TeV band (recall that their quiescent IC spectral hump peaks already at MeV or, sometimes, GeV energies), but they suffer both intrinsic and external absorption at these energies. Intrinsically, TeV emission produced inside the BLR will be absorbed, rendering invisible TeV flaring regions too close to the central engine. Furthermore, because FSRQs (unlike BL Lacs) are distributed in the Universe preferentially at higher redshift (Ackermann et al. 2015; Ajello et al. 2020), their TeV gamma-rays may be absorbed by the extragalactic background light while en route toward Earth. Owing to these combined effects, only a handful of FSRQs have even been detected at TeV energies (9 at the time of writing, Wakely & Horan 2008). Of these, quiescent emission has only been seen from one object, PKS 1510–089 (MAGIC Collaboration 2018). The rest are detected exclusively in high or flaring states (e.g. 3C 279, MAGIC Collaboration 2008; PKS 1222+216, Aleksić et al. 2011; PKS 1441+25, Abeysekara et al. 2015; QSO B0218+257, Sitarek et al. 2015; Ahnen et al. 2016; PKS 0736+017, H. E. S. S. Collaboration 2020; QSO B1420+326, MAGIC Collaboration 2021). Even during flares, the gamma-ray flux is rarely sufficient to provide detailed temporal information for the total luminosity itself, much less for the spectrum.

Considering mostly time-integrated spectra, what is generally seen for FSRQs in the TeV band is a much steeper spectrum, even after correcting for absorption by the extragalactic background light, than in the *Fermi* range. TeV spectral indices are almost always greater than 2.5 and commonly larger than 3: for 3C 279, $\Gamma \simeq 4.2$ (H. E. S. S. Collaboration 2019); for PKS 1510–089, $\Gamma \simeq 2.5$ (Aleksić et al. 2014), $\Gamma \simeq 3.2$ and 4.3 (Ahnen et al. 2017), $\Gamma \simeq 2.9$ and 3.4 (Zacharias et al. 2017), and $\Gamma \simeq 3.3$ (MAGIC Collaboration 2018); for PKS 1222+216, $\Gamma \simeq 2.7$ (Aleksić et al. 2011); for PKS 1441+25, $\Gamma \simeq 3.4$ (Abeysekara et al. 2015); for QSO B0218+257, $\Gamma \simeq 2.4$ (Ahnen et al. 2016); and for QSO B1420+326, $\Gamma \simeq 2.9$ (MAGIC Collaboration 2021). In the IC(HDR) scenario, these steep spectral indices suggest that reconnection proceeds in a regime, unlike that probed in detail in this study, where its intrinsic particle acceleration index is steeper (e.g. with a strong guide field: Werner & Uzdensky 2017). Even in this case, Klein–Nishina radiative physics should still harden the spectrum in the decaying part of a flare (Section 4.5). However, such detailed temporal behaviour has not

yet been accessible to TeV instruments. From this point of view, the results of our model remain, for the moment, predictions. The coming online of the Cherenkov Telescope Array (CTA) over the next few years will provide increased sensitivity in the TeV band (Cherenkov Telescope Array Consortium et al. 2019), enabling enhanced temporal resolution and, hence, a more thorough probing of the IC(HDR) scenario.

We note that BL Lacs are typically much brighter than FSRQs at TeV energies, and thus many more IACT observations of BL Lacs exist – and typically with much greater temporal resolution – than of FSRQs (Wakely & Horan 2008). However, without strong external photon fields supplied by the BLR or the HDR, the source of seed photons for reconnection in BL Lacs is far less constrained. If these are the synchrotron photons produced during reconnection (a single-zone synchrotron self-Compton, or SSC, setup, cf. Maraschi, Ghisellini & Celotti 1992; Bloom & Marscher 1996), a much more detailed seed photon model – taking into account multichromaticity, time-dependence, anisotropy, and spatial inhomogeneity – needs to be folded in to the simulations in order to make robust predictions. However, if the seed photons impinge on the putative reconnection region from elsewhere in the jet (a multizone SSC paradigm), then our static seed photon model is more appropriate, but it would still potentially need to be generalized to the non-monochromatic case. We note that in the famous minute-scale flare of PKS 2155–304 presented by Aharonian et al. (2007), no evidence of spectral variability was found. This is consistent with our reconnection model, however the photon indices, even after correcting for the background absorption (e.g. Aharonian et al. 2009) are steeper than one would expect on the decaying side of the flare, wherein Klein–Nishina and pair-production physics should harden the spectrum closer to the universal shape of Section 4.5. Ultimately, not enough is known about the seed photons to draw firm conclusions, however.

We close this discussion by examining the potential pair yield of Klein–Nishina reconnection in the IC(BLR) and IC(HDR) scenarios. Suppose that the jet is launched with an electron–proton composition (no positrons). Let us also assume that the jet evolves so as to be moderately magnetized in its rest frame, $B_0^2/4\pi n_0 m_p c^2 \sim 1-10$, at the parsec scale (cf. Giannios 2013; Giannios & Uzdensky 2019; Mehlhaff et al. 2021), which is near the transition point where the HDR overtakes the BLR as the dominant seed photon source (Nalewajko et al. 2012; Mehlhaff et al. 2021). The cold electron magnetization is then $\sigma_{c,0} \sim (m_p/m_e)(B_0^2/4\pi n_0 m_p c^2) \sim 2\times 10^{3-4}$. In our earlier work (Mehlhaff et al. 2021), we estimate (in the jet's rest frame) $\gamma_{\mathrm{KN,\,BLR}} \sim 300$ and $\gamma_{\mathrm{KN,\,HDR}} \sim 1\times 10^4$. This means that the pair yield control parameter (Section 5) is $\gamma_{\mathrm{pp,\,max}}/4\sigma_{c,0} \simeq 30\gamma_{\mathrm{KN}}/4\sigma_{c,0} \sim 0.1-1$ for reconnection illuminated by the BLR and 4–40 for HDR irradiation. This control parameter only becomes small (the regime of high pair yield) in the IC(BLR) scenario, which is, incidentally, also the scenario for which available flaring observations, as discussed in this section, best corroborate a Klein–Nishina reconnection model. If pair production is predominantly decided by leptonic physics as unveiled by our pair-plasma simulations, then the present estimates suggest that reconnection functions as an *in situ* source of antimatter in FSRQ jets strongly illuminated by the BLR, possibly creating more than 1 positron per electron and effectively transforming any initially electron–proton plasma into a plasma with a prominent positron component. This is an important result in blazar studies because the jet composition is notoriously difficult to ascertain observationally (Madejski & Sikora 2016); if it is true, it could mean that BLR-illuminated FSRQ jets generically carry a strong pair-plasma component downstream of the GeV emission zone.

6.1.3 *Synopsis of FSRQ discussion*

In summary, the bright Doppler-boosted emission from blazars highlights the physics of particle acceleration in their jets. In particular, for the FSRQ sources, GeV and TeV observations function as respective probes of scenarios where magnetic reconnection is coupled, through Klein–Nishina and pair-production physics, to soft seed photons produced by the BLR and the HDR. Observations of FSRQ flares in the GeV band are in broad agreement with the spectral-temporal signatures of Klein–Nishina reconnection expected from this study (Section 4.5), constituting potential evidence that this type of reconnection occurs within the BLR of FSRQs. The expected pair yield (Section 5) of IC(BLR)-coupled reconnection could also contribute significantly to the antimatter content of FSRQ jets far away from their central engines. The IC(HDR) scenario cannot be probed at the same level of detail as the IC(BLR) case with the current generation of IACTs, a limitation which the CTA will help to overcome. Finally, while Klein–Nishina reconnection could also operate in BL Lacs, there are a lot more uncertainties concerning the seed photons in those sources, pushing detailed observational comparisons into the domain of future work.

6.2 Black hole accretion disc coronae

6.2.1 *Background information*

In our previous work, Mehlhaff et al. (2021), we made a case for Klein–Nishina reconnection operating in the coronae of accreting BHXRBs in their high/soft states. Conducting basic estimates, we showed that, for the case where an underlying optically thick, geometrically thin accretion disc (Shakura & Sunyaev 1973) supplies

$$\epsilon_{\mathrm{disc}} \sim 1\,\mathrm{keV} \tag{58}$$

seed photons to reconnection in a highly magnetized coronal plasma, the radiative scale hierarchy is $\gamma_{\mathrm{cool}} < \gamma_{\mathrm{KN}} < \gamma_{\mathrm{rad,\,T}} < \gamma_{\mathrm{max}}$ [equation (50)], pushing reconnection into the Klein–Nishina regime. Due to the intense accretion disc radiation bath, absorption fiducially kicks in for gamma-rays above the threshold energy,

$$\epsilon_{\mathrm{th,disc}} = m_e^2 c^4/\epsilon_{\mathrm{disc}} = 260\,\mathrm{MeV}\,. \tag{59}$$

Meanwhile, the typical photon energy emitted by a γ_{KN} particle is $\gamma_{\mathrm{KN,disc}} m_e c^2/2 = \epsilon_{\mathrm{th,disc}}/8 = 30\,\mathrm{MeV}$, where $\gamma_{\mathrm{KN,\,disc}} \sim 100$.

To our knowledge, such energies have only been detected in the high/soft state of an accreting BHXRB – for which their origin in an ejected jet is not expected – in one object: Cyg X-1. This was during an approximately 100-Ms exposure by the *Fermi*-LAT, presented by Zdziarski et al. (2017), who report the detection of gamma-rays up to a cut-off energy of about $20-40\,\mathrm{MeV}$, somewhat below (59). As pointed out in Mehlhaff et al. (2021), this cut-off may be consistent with gamma-ray absorption because the coronal region is likely highly radiatively compact (which translates into a high fiducial pair-production optical depth, $\tau_{\gamma\gamma} \gg 1$), meaning that absorption is still prominent at energies below $\epsilon_{\mathrm{th,disc}} \sim 300\,\mathrm{MeV}$ by seed photons in the exponential tail of the disc spectrum.

6.2.2 *Connection to this work*

Given the long exposure time necessary for the gamma-ray detection (Zdziarski et al. 2017), the prospects for examining temporal signatures of Klein–Nishina reconnection in high/soft BHXRBs at the energy scales near $\gamma_{\mathrm{KN,\,disc}} m_e c^2$, where Klein–Nishina effects are likely most prominent, are not promising. At lower, X-ray energies,

millisecond variability was detected from the high/soft state of Cyg X-1 by Gierliński & Zdziarski (2003). Unlike the case of the IC(BLR) and IC(HDR) scenarios discussed in the preceding section, these X-ray observations exhibit a strong harder-when-brighter correlation between the flaring flux and the spectral shape. This could hint at the importance of synchrotron cooling in this system. Alternatively, due again to the high radiative compactness, the radiative cooling time-scale may be so short that, even in the presence of Klein–Nishina and pair-production effects, the reconnection-energized particles cool down instantaneously on the observationally resolved time-scales, masking the radiatively stabilized (on presumably faster time-scales; Section 4.5) spectrum.

Using our previous estimate of the coronal magnetization (Mehlhaff et al. 2021), $\sigma_{\rm c,0} \sim 10^{2\text{–}4}$, we can infer a pair yield control parameter of $\gamma_{\rm pp,\,max}/4\sigma_{\rm c,0} \sim 30\gamma_{\rm KN,\,disc}/4\sigma_{\rm c,0} \sim 0.08\text{–}8$, indicating Klein–Nishina reconnection as a potentially important source of electron–positron pairs in the coronae of BHXRBs in their high/soft states. However, due to the high radiative compactness, another likely source of pair production is the collision of disc-Comptonized photons with each other (Beloborodov 2017). This interaction occurs predominantly among photons both with energies $< \epsilon_{\rm th,disc} = m_e^2c^4/\epsilon_{\rm disc}$, since those with energies beyond $\epsilon_{\rm th,\,disc}$ are most likely to be absorbed by the much denser disc-supplied radiation field. The total pair yield is then the sum of the contributions from both channels.

6.2.3 *Synopsis of BHXRB discussion*

To summarize, the case for coronal Klein–Nishina reconnection in the high/soft states of BHXRBs is excellent on theoretical grounds (Mehlhaff et al. 2021). It is likely, for example, that reconnection in this regime contributes to the pair content in these objects (Section 5). However, given that even in an exceptionally bright source, Cyg X-1, the gamma-ray flux is far too low to probe corona-scale reconnection-powered flaring variability, firm connections with the characteristic temporal signatures of Klein–Nishina reconnection (Section 4.5) will likely remain out of reach for the foreseeable future.

6.3 The M87* magnetosphere

6.3.1 *Background information*

The M87 galaxy has been monitored in TeV gamma-rays for nearly 20 yr (e.g. Aharonian et al. 2006; MAGIC Collaboration 2020; EHT MWL Science Working Group et al. 2021). This includes three major flaring periods, one in 2006 (Aharonian et al. 2006), one in 2008 (MAGIC Collaboration 2008; Acciari et al. 2009), and one in 2010 (Abramowski et al. 2012; Aliu et al. 2012), that exhibited rapid, one-day variability time-scales – of order the light crossing time of the galaxy's central supermassive black hole, M87*. For two of these TeV-loud periods, observations at longer, more spatially resolved wavelengths revealed (nearly) contemporaneous flux enhancements from the galactic core (Acciari et al. 2009; Abramowski et al. 2012). Thus, variability considerations combined with the multiwavelength context suggest M87* itself (more precisely, its immediate plasma environment) as a viable site of TeV emission.

Direct polarized imaging of M87* suggests that accretion proceeds in a so-called magnetically arrested (MAD) state (Event Horizon Telescope Collaboration 2021), characterized, as revealed in large part by magnetohydrodynamics simulations, by quasi-periodic cycles of gradual accumulation of magnetic flux onto the black hole, eventual flux saturation, and finally abrupt and violent flux expulsion (e.g. Igumenshchev 2008; Tchekhovskoy, Narayan & McKinney 2011; Avara, McKinney & Reynolds 2016; Ripperda et al. 2020, 2022; Chashkina, Bromberg & Levinson 2021; Porth et al. 2021; Mizuno 2022; Scepi et al. 2022). These expulsion events are mediated in the black-hole magnetosphere by large-scale reconnecting current sheets (Ripperda et al. 2022). The current sheets are irradiated by the larger-scale accretion flow, which provides a target photon bath for reconnection-accelerated particles to Comptonize up to the observed TeV energies (Hakobyan et al. 2023b). If particle acceleration is efficient enough, the tail of the Comptonized radiation spectrum falls above pair-production threshold with the accretion flow seed photons, triggering potentially copious pair creation (Crinquand et al. 2021, 2022). Thus, reconnection in the M87* magnetosphere is: (1) a potential source of the observed TeV emission, and (2) likely coupled to the same radiative physics as treated in this study.

6.3.2 *Connection to this work*

We first demonstrate that, in line with point (2) above, the Klein–Nishina hierarchy (50) is likely realized in the M87* magnetosphere, pointing to the importance of Klein–Nishina effects on reconnection there. We adopt a black hole mass $M_{\rm BH} = 6 \times 10^9 M_\odot$ (Gebhardt et al. 2011; Event Horizon Telescope Collaboration 2019) and horizon-scale reconnecting magnetic field strength $B_{\rm BH} = 100\,{\rm G}$ (Yao et al. 2021; Ripperda et al. 2022). We also use the seed photon radiation field calculated by Yao et al. (2021) that is produced in a MAD accretion state and illuminates the inner black hole magnetosphere with overall energy density $U_{\rm BH} = 10\,{\rm erg\,cm^{-3}}$ and characteristic photon energy $\epsilon_{\rm BH} \sim 10^{-2}\,{\rm eV}$. Then, using equation (7), we arrive at

$$\gamma_{\rm KN,BH} = \frac{m_e c^2}{4\epsilon_{\rm BH}} \sim 1 \times 10^7\,, \tag{60}$$

while, via equation (48), we estimate

$$\gamma_{\rm rad,T}^{\rm (BH)} = \sqrt{\frac{0.3 e B_{\rm BH}}{4\sigma_{\rm T} U_{\rm BH}}} \sim 2 \times 10^7\,. \tag{61}$$

In addition, adopting a fiducial reconnection current sheet length of order the gravitational radius, $r_g = GM_{\rm BH}/c^2 \simeq 9 \times 10^{14}\,{\rm cm}$, gives, by equation (38),

$$\gamma_{\rm max,BH} = \frac{0.1 e B_{\rm BH} r_g}{m_e c^2} \sim 5 \times 10^{12} \tag{62}$$

and, through $\gamma_{\rm rad,T}^2 = \gamma_{\rm max}\gamma_{\rm cool}$ [equation (49)],

$$\gamma_{\rm cool,BH} \sim 100\,, \tag{63}$$

which can be used, via equation (51), to evaluate

$$\tau_{\gamma\gamma,\rm BH} = \frac{3\gamma_{\rm KN,BH}}{5\gamma_{\rm cool,BH}} \sim 7 \times 10^4\,. \tag{64}$$

Thus, the Klein–Nishina scale hierarchy (50), $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, likely holds (though $\gamma_{\rm KN,\,BH}$ in these estimates somewhat encroaches on $\gamma_{\rm rad,T}^{\rm (BH)}$ from below).

Although the above estimates suggest a coupling of reconnection to Klein–Nishina physics, synchrotron losses, unlike in our simulations, are non-negligible in the M87* magnetosphere. To show this, we note that the synchrotron power radiated per particle is

$$P_{\rm syn}(\gamma) = 2\sigma_{\rm T} c \beta^2 \gamma^2 U_{\rm B} \sin^2\alpha\,, \tag{65}$$

where $U_{\rm B} \equiv \boldsymbol{B}^2/8\pi$ is the local magnetic field energy density and α is the pitch angle between the radiating particle's velocity and

the local magnetic field. An approximate lower bound on the ratio of synchrotron radiation to IC radiation is then, via equations (65), (41), and (42),

$$\frac{P_{\rm syn}(\gamma)}{P_{\rm IC}(\gamma)} \geq \frac{P_{\rm syn}(\gamma)}{P_{\rm T}(\gamma)} = \frac{3U_{\rm B}\sin^2\alpha}{2U_{\rm BH}}\,, \tag{66}$$

where, in the first step, we ignore potential Klein–Nishina suppression of the IC cross-section. Thus, excepting special regions (such as reconnection X-points) where the magnetic field weakens ($\boldsymbol{B}^2 \ll B_{\rm BH}^2$) or the pitch angle shrinks ($\sin^2\alpha \ll 1$), synchrotron cooling typically exceeds IC losses by the factor $B_{\rm BH}^2/8\pi U_{\rm BH} \sim 40$.

Having recognized their importance, let us now reason about how strong synchrotron losses may change the observable aspects of reconnection as unveiled in this work. The fact that equation (65) is identical to the Thomson IC formula (41) but with $U_{\rm ph}$ replaced by $3U_{\rm B}\sin^2\alpha/2$ suggests that efficient synchrotron radiation and Thomson IC cooling may play dynamically similar roles. We might therefore expect, akin to our simulations of Thomson IC reconnection, radiation from the M87* magnetosphere to exhibit a much tighter coupling between spectral shape and total luminosity – that is, 'harder-when-brighter' – than when only Klein–Nishina IC cooling and pair production dominate the radiative physics (cf. Section 4.5). This would hold whether one observes the synchrotron or the IC radiation, as both probe the same underlying distribution of reconnection-energized particles. Such an argument for a prominent harder-when-brighter trend is, however, in mild tension with the few M87 TeV flares, which indicate this trend only mildly (Aliu et al. 2012) and not in every event (Acciari et al. 2010).

We next consider the potential ramifications of pair production between IC photons and the seed photons from the accretion flow. To provide a context for this discussion, we briefly summarize recent semi-analytic works, Kimura et al. (2022), Hakobyan et al. (2023b), and Chen et al. (2023), on the matter content of the putative magnetospheric M87* reconnection layer. Hakobyan et al. (2023b) conduct a detailed analysis of the possible radiation and pair-production channels in this system. They predict that the brightest radiation emerges through the synchrotron band and peaks between roughly 1 and 20 MeV. Though the synchrotron photons are not above pair-production threshold with the ambient accretion-flow photons, they are above threshold with themselves. There is, however, little chance for an individual synchrotron photon to be absorbed by another since the optical depth, $\tau_{\gamma\gamma}^{\rm (syn)} \sim 10^{-4}$, presented by the synchrotron radiation is small. Thus, whereas the high optical thickness furnished by the accretion-flow light causes nearly all of the above-threshold IC photons to be absorbed close to the reconnection current sheet, covering it with a thin pair coat, only a small fraction of the synchrotron radiation is absorbed, leading to diffuse pair production throughout the magnetosphere. The resulting pairs feed the reconnection layer with a highly magnetized, $\sigma_{\rm c,0} \sim 5 \times 10^7$, plasma. The pictures presented by Kimura et al. (2022) and Chen et al. (2023) are similar, but those authors predict more copious synchrotron–synchrotron pair production, yielding the reduced magnetizations, $\sigma_{\rm c,0} \sim 8.7 \times 10^4$ and $\sim 6 \times 10^4$, respectively.

In the context of our reconnection model, the diffuse synchrotron–synchrotron pair production predicted by Kimura et al. (2022), Hakobyan et al. (2023b), and Chen et al. (2023) supplies the background plasma on top of which pair production between IC photons and the radiation impinging from the accretion flow (hereafter, *IC pair production*) may add supplementary pairs localized near the current sheet. If we naively apply our findings from Section 5 on the reconnection-powered IC pair yield (temporarily ignoring potential changes due to synchrotron cooling), we infer that a substantial amplification of the background (synchrotron–synchrotron) pair density requires $4\sigma_{\rm c,0} \geq \gamma_{\rm pp,\,max} \simeq 30\gamma_{\rm KN} \sim 4 \times 10^8$. Comparing to $\sigma_{\rm c,0} \sim 5 \times 10^7$ from Hakobyan et al. (2023b), we estimate $\gamma_{\rm pp,\,max}/4\sigma_{\rm c,0} \simeq 2$, which is slightly too high to attain an order-unity IC pair yield according to this study (Section 5). One should keep in mind, however, that the uncertainties on these order-of-magnitude estimates are high, while even a reduction in the ratio $\gamma_{\rm pp,\,max}/4\sigma_{\rm c,0}$ by a factor of 4 is enough to bring the IC pair yield up to unity. However, if synchrotron–synchrotron pair production is more efficient, as predicted by Kimura et al. (2022) and Chen et al. (2023), then the inflowing plasma is probably not sufficiently magnetized to lead to appreciable IC pair production. In the event that a substantial IC pair yield is realized, it induces a transverse density gradient on the magnetic field lines forming the jet funnel, with a higher plasma load on the field lines that participate in reconnection near the jet walls. This would then be potentially important for jet-boundary interactions (e.g. Ripperda et al. 2020; Sironi, Rowan & Narayan 2021; El Mellah et al. 2022, 2023; Chow et al. 2023a, b), such as may power observed limb-brightening (Ly, Walker & Junor 2007; Kim et al. 2018; Walker et al. 2018).

In the preceding discussion, we simply apply the forecasted pair yield obtained from this study (Section 5) to the M87* case. We now anticipate, as we did earlier for radiative signatures, how strong synchrotron losses may alter the picture of pair production as revealed by this work. First, rapid synchrotron cooling disposes the IC pairs born into the upstream region to cool down much more before entering the reconnection layer, which may change them from a hot population to a merely warm or even cold one (thus reducing their contribution to the upstream plasma energy density; cf. Section 4.6). Second, the IC pair yield may be restricted since high-energy particles are likely to emit fewer pair-producing IC photons before radiating away their energy as synchrotron light. Interestingly, however, the global magnetospheric simulations of Crinquand et al. (2022), which include the same processes of IC emission and pair creation studied here, find that IC pair production is efficient enough to fuel the reconnection layer with plasma even when synchrotron losses are made as strong as numerically possible. This is likely connected to the fact that TeV emitting particles are accelerated near reconnection X-points (Hakobyan et al. 2023b) where their synchrotron losses are suppressed. Despite these arguments, then, previous numerical work hints that synchrotron radiation does not quench IC pair production.

6.3.3 *Synopsis of M87* discussion*

To summarize, the basic Klein–Nishina hierarchy (50) seems to be satisfied in the M87* magnetosphere. Nevertheless, synchrotron cooling is also an important radiative mechanism for reconnection in this system. Because we neglect this in our simulations, we cannot directly apply our results to the M87* case. We therefore, in this section, attempt to anticipate the modifications synchrotron cooling could induce on the radiative signatures and pair yield of Klein–Nishina reconnection as uncovered in this work (Sections 4.5 and 5). When formulating these expectations, we find that the resulting conclusions are not convincingly supported by recent studies. For example, we argue that synchrotron losses might induce a tighter correlation between spectral shape and total luminosity, making

the radiative signatures of reconnection more Thomson-like (with a more prominent harder-when-brighter trend). However, observational evidence for this is somewhat ambiguous, providing, at best, limited support (Aliu et al. 2012) and, at worst, mild tension (Acciari et al. 2010; Beilicke & VERITAS Collaboration 2012). We further argue that synchrotron cooling may shut down the IC pair yield, but this seems to be in conflict with first-principles global simulations (Crinquand et al. 2022). In short, comparing to observations (Acciari et al. 2010; Aliu et al. 2012; Beilicke & VERITAS Collaboration 2012) and simulations (Crinquand et al. 2022) suggests that our simple arguments do not satisfactorily generalize the results of this study to reconnection where efficient synchrotron cooling also comes into play. This creates fertile ground for future reconnection simulations to self-consistently incorporate synchrotron radiation on an equal footing with Klein–Nishina effects to illuminate how the two conspire together to shape the gamma-ray signatures and pair yield in the M87* magnetosphere.

6.4 Gamma-ray binaries

6.4.1 Background information

Gamma-ray binaries consist of a relativistic compact object (neutron star or black hole) and a massive stellar companion (generally of type O or Be) and are defined by a spectral energy density peaking, in the $\nu F(\nu)$ representation, above 1 MeV (e.g. Dubus 2013; Dubus et al. 2017). Of the handful (less than a dozen) of known gamma-ray binaries, only two are directly observed to host pulsars (Dubus 2013; Tam et al. 2018; Chernyakova et al. 2019; Chernyakova & Malyshev 2020). Nevertheless, several general observed features suggest that the compact object in these systems is generically a rotation-powered pulsar (Dubus 2006, 2013).

Adopting this view, two often-invoked gamma-ray emission sites are the pre- and post-shocked pulsar wind, where the shock in question interfaces between the winds of the pulsar and the massive companion (e.g. Kirk, Ball & Skjæraasen 1999; Ball & Kirk 2000; Ball & Dodd 2001; Sierpowska & Bednarek 2005; Cerutti et al. 2008; Sierpowska-Bartosik & Bednarek 2008; Khangulyan et al. 2012) and not, as would be the case in isolated pulsars, between the pulsar wind and the interstellar medium. As argued in the review by Dubus (2013), the fact that the high-energy ($0.1 - 10\,\mathrm{GeV}$) spectra of gamma-ray binaries are often similar in terms of slope and cut-off to isolated pulsars could hint at a similar emission mechanism between the two object classes, motivating an investigation of the unshocked pulsar wind, and perhaps even of the pulsar magnetosphere, as gamma-ray production sites in gamma-ray binaries. However, the GeV spectra also exhibit modulations on the binary orbital period, which would seem to disfavour the magnetosphere as the dominant emission zone, since it is insensitive to the orbit of the binary (Dubus 2013). Therefore, in this section, we consider the possibility that the unshocked pulsar wind significantly contributes to the observed high-energy gamma-rays in gamma-ray binaries (cf. Ball & Kirk 2000; Ball & Dodd 2001; Cerutti et al. 2008; Khangulyan et al. 2012).

The unshocked pulsar wind behaves exactly as that of an isolated pulsar except for one key difference: it is illuminated from beyond by the hot massive companion star. The temperature, $T_\star \sim 40\,000\,\mathrm{K}$, of the companion's surface produces a characteristic blackbody photon energy,

$$\epsilon_\star \sim 3k_\mathrm{B}T_\star \sim 10\,\mathrm{eV}\,, \tag{67}$$

and radiation energy density,

$$U_\star = \frac{\sigma_\mathrm{SB}T_\star^4}{c}\left(\frac{R_\star}{d}\right)^2 \sim 1\times10^3\,\mathrm{erg\,cm^{-3}}\,, \tag{68}$$

where $d = 0.1\,\mathrm{au}$ is the typical separation at periastron and $R_\star = 10\,\mathrm{R}_\odot$ is the stellar radius (Dubus 2013). Equation (67) implies a critical Klein–Nishina Lorentz factor in the pulsar wind of

$$\gamma_\mathrm{KN,w} \equiv \frac{m_e c^2}{4\epsilon_\star} \sim 1\times10^4 \tag{69}$$

and that pair production becomes possible above the threshold

$$\epsilon_\mathrm{th,w} = m_e^2c^4/\epsilon_\star = 30\,\mathrm{GeV}\,. \tag{70}$$

The typical photon energy emitted by particles with $\gamma = \gamma_\mathrm{KN,\,w}$ is then $\gamma_\mathrm{KN,w}m_ec^2/2 = \epsilon_\mathrm{th,w}/8 \sim 3\,\mathrm{GeV}$.

6.4.2 Connection to this work

Let us consider the effect that illumination by the companion may have on the pulsar wind. Here, we adopt the theoretical picture (Coroniti 1990; Michel 1994; Bogovalov 1999; Lyubarsky & Kirk 2001; Kirk & Skjæraasen 2003; Lyubarsky 2003; Kirk, Lyubarsky & Petri 2009), brought into sharper focus by recent first-principles kinetic simulations (Cerutti & Philippov 2017; Philippov & Spitkovsky 2018; Cerutti et al. 2020), that this wind is not purely cold, but is instead *striped* – laced with a large-scale reconnecting current sheet that expands radially outward while undulating about the pulsar's rotational equator. The angular excursions about the equator approximately equal the obliquity angle between the magnetic and spin pulsar axes. Reconnection converts the outgoing wind Poynting flux into a combination of bulk acceleration and NTPA.

An isolated pulsar is not bathed in the intense light of a companion star, and so cooling of accelerated wind particles remains dominated by synchrotron losses. In the present case, however, cooling via IC scattering of the intense radiation bath (68) likely outpaces synchrotron cooling once the local magnetic field energy density falls below $U_\star$. This occurs at a critical magnetic field strength,

$$B_\mathrm{IC} \equiv \sqrt{8\pi U_\star} \sim 200\,\mathrm{G}\,. \tag{71}$$

For a pulsar rotational period $P \sim 100\,\mathrm{ms}$ (characteristic of the two confirmed pulsars in gamma-ray binaries: PSR J2032 + 4127, Abdo et al. 2009a; and PSR B1259−63, Johnston et al. 1992) and a surface magnetic field $B_\mathrm{surf} = 10^{12}\,\mathrm{G}$, the magnetic field is diluted to $B_\mathrm{LC} \simeq B_\mathrm{surf}(R_\mathrm{psr}/R_\mathrm{LC})^3 \sim 9\times10^3\,\mathrm{G}$, at the light cylinder, $R_\mathrm{LC} = cP/2\pi$, where $R_\mathrm{psr} = 10\,\mathrm{km}$ is the assumed pulsar radius. Beyond the light cylinder, the magnetic field falls off slower, as R_LC/R where R is the cylindrical radius measured from the pulsar's spin axis, and, hence, even for the strong surface field $B_\mathrm{psr} = 10^{12}\,\mathrm{G}$, the striped wind's radiative losses become IC dominated at $R/R_\mathrm{LC} \simeq (B_\mathrm{surf}/B_\mathrm{IC})(R_\mathrm{psr}/R_\mathrm{LC})^3 \sim 60$. This is far before the shock with the companion's wind, which is expected to occur on scales $R \sim d \sim 10^3R_\mathrm{LC}$ (Dubus 2013). Moreover, $R \sim 60R_\mathrm{LC}$ is also before the point where the pulsar wind's electromagnetic flux is expected to be fully dissipated, which kinetic simulations anticipate at R/R_LC roughly between 10^2 and 10^4 (Cerutti et al. 2020). Taken together, these estimates suggest that *most of the pulsar wind's dissipation takes place before the shock with the companion's wind, but after the critical radius where IC losses surpass synchrotron losses*.

Not only does most of the pulsar wind dissipate through reconnection subject to strong IC cooling, but, as we show now, reconnection may occur in the Klein–Nishina regime of this study, attaining the

critical scale hierarchy (50), $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$. Assuming $B_{\rm LC} = 9 \times 10^3$ G, $P = 100$ ms, and that the reconnecting magnetic field strength, $B_{\rm w}$, is $B_{\rm w} = B_{\rm LC} R_{\rm LC}/R$, as well as identifying the length, L, of the reconnection layer with the local radius, R, in the pulsar wind, we have, by equation (38),

$$\gamma_{\rm max,w} \equiv \frac{0.1 e B_{\rm w} R}{m_e c^2} \sim 3 \times 10^8 \,, \tag{72}$$

which is independent of R. Similarly, we have, by equation (48),

$$\gamma^{\rm (w)}_{\rm rad,T} \equiv \sqrt{\frac{0.3 e B_{\rm w}}{4\sigma_{\rm T} U_\star}} \sim 3 \times 10^6 \left(\frac{R}{60 R_{\rm LC}}\right)^{-1/2} . \tag{73}$$

We then can use $\gamma^2_{\rm rad,T} = \gamma_{\rm cool}\gamma_{\rm max}$ [equation (49)], to estimate

$$\gamma_{\rm cool,w} \sim 3 \times 10^4 \left(\frac{R}{60 R_{\rm LC}}\right)^{-1} . \tag{74}$$

Finally, using equation (51), we estimate the pair-production optical depth as

$$\tau_{\gamma\gamma,\rm w} \equiv \frac{3\gamma_{\rm KN,w}}{5\gamma_{\rm cool,w}} \sim 0.3 \left(\frac{R}{60 R_{\rm LC}}\right) . \tag{75}$$

Note that we have normalized R to the critical radius where IC losses overtake synchrotron losses. The pulsar wind extends much farther than this, which pushes $\tau_{\gamma\gamma,\rm w}$ above unity at the largest radii. The geometry of the striped wind is somewhat peculiar, however, in that the spacing between stripes is $\sim R_{\rm LC} \ll R$ and, hence, even if the overall current sheet length is optically thick to pair production, the spacing between current sheets is thin, such that photons emitted in one stripe may be absorbed inside another.

Let us now examine what the implications of Klein–Nishina reconnection are on the pre-shocked pulsar wind of gamma-ray binaries. We discuss first the potential effect of pair production on the wind. The cold magnetization at the light cylinder is

$$\sigma^{\rm (w)}_{\rm c,0} = \frac{B^2_{\rm LC}}{4\pi\kappa n_{\rm GJ} m_e c^2} = \frac{e P B_{\rm LC}}{4\pi\kappa m_e c} = 1 \times 10^5 \left(\frac{B_{\rm LC}}{9 \times 10^3\,{\rm G}}\right)\left(\frac{\kappa}{10^4}\right)^{-1} , \tag{76}$$

where $n_{\rm GJ} \equiv B_{\rm LC}/ecP$ is the Goldreich–Julian number density and $\kappa = 10^4$ is a fiducial pair multiplicity (e.g. Timokhin & Harding 2015, 2019). This magnetization is 'frozen-in' at the light cylinder in the sense that, because $B_{\rm w} \propto 1/R$ and $n_0 \propto 1/R^2$ beyond the light cylinder, $\sigma^{\rm (w)}_{\rm c,0}$ remains constant (modulo pair production) in the unreconnected plasma of the wind throughout its expansion. The pair yield control parameter in the pre-shocked wind is then $\gamma_{\rm pp,max}/4\sigma^{\rm (w)}_{\rm c,0} \simeq 30\gamma_{\rm KN,w}/4\sigma^{\rm (w)}_{\rm c,0} \sim 0.7$, which is in the regime of order-unity pair yield. Furthermore, if the pulsar magnetosphere fails to launch the pulsar wind with $\kappa \sim 10^4$, but instead with a lower multiplicity, the magnetization $\sigma^{\rm (w)}_{\rm c,0}$ increases, leading to potentially copious *in situ* pair production in the expanding wind. Using our results from Section 5, we infer a critical self-regulated magnetization of $\sigma^{(\gamma\gamma)}_{\rm c} \simeq \gamma_{\rm pp,max} \simeq 30\gamma_{\rm KN,w} \sim 4 \times 10^5$. If the plasma is injected with a higher magnetization than this (e.g. by virtue of an underdense multiplicity), Klein–Nishina pair production fills in the plasma deficit, pulling up the multiplicity toward $\kappa = eP B_{\rm LC}/4\pi m_e c \sigma^{(\gamma\gamma)}_{\rm c} \sim 3 \times 10^3$. This critical multiplicity depends solely on the pulsar properties and those of the ambient radiation field.

Let us close by examining prospects for observing signatures of Klein–Nishina reconnection in gamma-ray binaries. The temporal observables uncovered in Section 4.5 likely manifest themselves most clearly during transient flares. Of the known gamma-ray binaries, PSR B1259−63 is known to flare once per 3.4-yr orbit (e.g. Abdo et al. 2011; Tam et al. 2011, 2018; Caliandro et al. 2015; Chang et al. 2021). However, this binary is far less compact than others, with $d \sim 0.9$ au $\gg 0.1$ au, even at periastron. Moreover, the flares occur significantly later in the orbit than the time of periastron. Such wide separations dilute the radiation field from the stellar companion, tending to move reconnection out of the Klein–Nishina regime. The binaries LS 5039, LS I +61°303, and 1FGL J1018.6−5856 are all much more compact ($d \sim 0.1$ au at periastron; Dubus 2013), but, instead of sudden flares, GeV gamma-ray observations reveal smooth modulations on each binary's orbital period (for LS 5039, Abdo et al. 2009c; for LS I +61°303, Abdo et al. 2009b; for 1FGL J1018.6−5856, Fermi LAT Collaboration 2012). Such modulations likely probe quasistatic changes to the unshocked pulsar wind, including its orbit-dependent illumination by the companion, rather than variability intrinsic to reconnection (a situation anticipated in the concluding remarks of Section 4.5). In order to interpret these modulations, one needs to go beyond the flaring-centric treatment of observable signatures adopted in this work and conduct explicit global modelling of the steady-state wind.

6.4.3 *Synopsis of gamma-ray binary discussion*

To summarize, we have shown that Klein–Nishina reconnection likely takes place in short-period gamma-ray binaries if the compact object in the binary is a pulsar. In that case, reconnection occurs in the pre-shocked striped pulsar wind and is immersed in a bright seed photon bath supplied by the companion star. As a result, the reconnecting stripes produce a minimum pair-plasma density, setting a multiplicity/density floor in the wind (even if it is launched underdense from the pulsar). Observational support for this scenario is somewhat limited, as the few observed bright flares occur in configurations where the stellar companion is too far removed to supply a sufficiently dense radiation bath for Klein–Nishina reconnection. More compact systems, on the other hand, do not exhibit rapid flares, but rather smooth modulations to their GeV signal over the entire orbit. Detailed predictions of this signal require global modelling, which we leave to a future study.

6.5 Overall discussion summary

We supply a graphic recapitulation of the discussion in this section in Table 4. In all four types of systems – FSRQs, (high/soft-states of) black hole accretion disc coronae, the M87* magnetosphere, and gamma-ray binaries – a strong case can be made that the basic Klein–Nishina reconnection scale hierarchy (50), $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,\,T} < \gamma_{\rm max}$, is realized. However, only for the FSRQs [particularly the IC(BLR) scenario] do suitable observations exist for comparing with the expected temporal signatures of Klein–Nishina reconnection. For these objects, we find that *Fermi*-LAT data are, on the whole, consistent with the picture of a reconnection-powered flaring spectral energy density stabilized by pair production (Section 4.5). For the other object classes, more detailed observational connections are inhibited by: current instrumental sensitivity for the FSRQ IC(HDR) scenario (see Section 6.1; the coming online of the CTA may soon alleviate this); intrinsically dim sources for the accretion disc coronae of BHXRBs (see Section 6.2; this difficulty will probably not be overcome); and modelling details for the M87* magnetosphere and gamma-ray binaries (see Sections 6.3 and 6.4; the most pressing needs here are to account for synchrotron cooling in the case of M87* and for the global geometry of gamma-ray binaries).

7 CONCLUSIONS

In this work, we construct (Section 3) a numerical model of *Klein–Nishina reconnection*. We consider a reconnection system immersed in a background radiation bath that is static, homogeneous, isotropic, and monochromatic, characterized solely by its total energy density $U_{\rm ph}$ and individual photon energy $\epsilon_{\rm ph}$ [equation (4)]. We restrict to a 2D reconnection setup, a consequence of the large box sizes needed to maintain an adequate separation between the many scales in the problem, several of which are introduced by the QED physics. In particular, we need to satisfy the critical hierarchy of Lorentz factor energy scales (50), $\gamma_{\rm cool} \ll \gamma_{\rm KN} \ll \gamma_{\rm rad,\,T} \ll \gamma_{\rm max}$. This hierarchy, which is apparently satisfied in a variety of astrophysical systems (Section 6; Mehlhaff et al. 2021), permits: (1) efficient IC radiative losses on time-scales much shorter than the light crossing time, L/c; (2) copious particle acceleration above the energy, $\sim\gamma_{\rm KN}$, where particles emit photons above pair-production threshold; and (3) a high optical depth, $\tau_{\gamma\gamma} \sim \gamma_{\rm KN}/\gamma_{\rm cool} \gg 1$, such that nearly all of the above-threshold radiation is absorbed inside the system.

To model Klein–Nishina effects, we use a QED-enabled version of the ZELTRON PIC code, the development of which is presented in Section 2. We employ two main numerical optimizations to speed up the Monte Carlo IC scattering and pair production (Section 2.2.3): the null collision method and an approximation to the IC cross-section. Of these two, the former is widely used, while, prior to this work, the latter has, to our knowledge, only been implemented in GRZELTRON (see the supporting information of Crinquand et al. 2020), and we supply here the first exhaustive description.

We deploy the capabilities developed in Section 2 to run simulations of Klein–Nishina reconnection, which we present in Sections 4 and 5. These include various control runs – one non-radiative, one Thomson-radiative, and one Klein–Nishina-radiative with pair production artificially suppressed – to elicit the unique properties induced by the novel QED physics. We draw the following main conclusions for Klein–Nishina reconnection:

(i) From Sections 4.1 and 4.2: The added Klein–Nishina and pair-production physics does not substantially change the large-scale qualitative aspects of reconnection: the reconnection rate and hierarchical plasmoid chain remain essentially unchanged.

(ii) From Section 4.3: Reconnection-powered NTPA remains efficient but is somewhat inhibited by radiative cooling. On time-average, the reconnection-energized plasma is colder, and the particle energy distribution's extended non-thermal tail steeper, compared to the non-radiative regime, but the plasma is not as cold, nor the distribution as steep, as in the Thomson radiative regime.

(iii) From Section 4.4: The time-averaged IC emission spectrum is steeper than in the non-radiative regime but exhibits little change in slope from the Thomson-cooled case. This owes to Klein–Nishina effects suppressing the IC cross-section, and, hence, largely cancelling the hardening of the underlying particle energy distribution with respect to the Thomson case (cf. Moderski et al. 2005).

(iv) From Section 4.5: The intrinsic IC emission peaks above pair-production threshold, meaning that most of the initially radiated energy is reprocessed (by radiative cooling and pair production) to longer wavelengths before escaping the system. This leads, despite similarities in time-averaged spectra, to profound differences in the time-dependent signatures of reconnection between the Klein–Nishina and Thomson radiative regimes. In the latter case, the observed spectrum exhibits a tight correlation between total luminosity and prominence of the high-energy non-thermal tail (i.e. harder-when-brighter). In the Klein–Nishina case, however, the spectral shape becomes virtually independent of total luminosity, including in the decaying phase after reconnection-powered particle energization has ceased.

(v) From Section 4.5: When particle energization is shut off, Klein–Nishina IC cooling and pair production conspire to produce a universal spectral shape, with the particle energy distribution scaling as $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-2}$ for $\gamma < \gamma_{\rm KN}$ and IC emission spectrum as $\epsilon F_{\rm IC}(\epsilon) \propto \epsilon^{1/2}$ for $\epsilon < \gamma_{\rm KN} m_e c^2$.

(vi) From Section 4.6: The upstream plasma enthalpy – and, hence, the upstream hot magnetization – is not sufficiently loaded by newborn pairs to trigger the limit cycles sketched by Mehlhaff et al. (2021), according to the pair-loading efficiency requirement derived by those authors. This conclusion, however, applies only when the newborn upstream pairs are hot and tenuous, contributing significantly to the pressure of the reconnection inflow plasma but not to its number density. In regimes (suggested to exist by our findings concerning the pair yield below) where pair production loads the upstream plasma number density – and, thus, the cold magnetization in addition to the hot magnetization – the possibility of limit cycles remains an open issue.

(vii) From Section 5: The pair yield (per reconnection-processed lepton) follows an empirical exponential decay law, equation (53), in the parameter, $\gamma_{\rm pp,\,max}/4\sigma_{\rm c,0} \simeq 30\gamma_{\rm KN}/4\sigma_{\rm c,0}$. When $\gamma_{\rm pp,\,max} \lesssim 4\sigma_{\rm c,0}$, reconnection-powered NTPA cuts off at $\gamma \sim \gamma_{\rm X} \equiv 4\sigma_{\rm c,0}$, far beyond the energy, $\gamma_{\rm pp,\,max}$, where typical Comptonized photons lie at peak pair-production cross-section with the seed photons, enabling an order-unity pair yield. We speculate that the exponential formula breaks down in the limit, $4\sigma_{\rm c,0} \gg \gamma_{\rm pp,\,max}$ (not probed by our simulations), where reconnection would instead produce copious pairs. In such a case, the population of newborn pairs would no longer be hot and tenuous as in Section 4.6, but rather hot and abundant.

(viii) From Section 6: The Klein–Nishina reconnection scale hierarchy (50) is likely satisfied in at least four classes of astrophysical systems: FSRQs, high/soft states of BHXRBs, the magnetosphere of M87*, and gamma-ray binaries. Of note are FSRQs, where the often-observed spectral stability during GeV flaring states could be due to Klein–Nishina effects as discussed (and summarized above) in Section 4.5. For the other object classes, further modelling or instrumental development will enable firmer observational connections. A more detailed summary of Section 6 can be found in Section 6.5 and Table 4.

A few broad implications of Section 4.5's results merit additional discussion. The spectral stability properties in that section stem from the similar shape of the Klein—Nishina reconnection IC spectrum to that produced when an initial population of particles is allowed to passively cool through IC emission and pair production. This introduces a fundamental degree of degeneracy in efforts to leverage astrophysical observations to learn about plasma physical particle acceleration processes. Namely, any Klein–Nishina-coupled particle accelerator with an IC spectrum resembling that to which the radiative physics inevitably relaxes (i.e. after the accelerator is turned off) may yield the same spectral stability as reconnection. However, this degeneracy is, thankfully, incomplete, because acceleration processes with softer intrinsic spectra would instead likely yield an asymmetry between the rising side of a gamma-ray flare (where the steep particle acceleration spectrum would be visible) and the decaying phase (where the shallower Klein–Nishina-decaying spectrum would probably dominate).

In the face of such degeneracy, observations of temporally resolved spectra provide precious information. Already in this study, such spectra provide crucial and obvious distinctions between Thomson and Klein–Nishina reconnection – two regimes whose time-averaged

spectra have practically identical photon indices – and, in the case of FSRQs (Section 6), supply compelling evidence for the operation of the latter over the former. In the future, time-resolved spectra may be necessary to distinguish among different possible Klein–Nishina-coupled particle acceleration mechanisms. Hence, this work strongly motivates collection of time-resolved (versus simply time-averaged) spectra whenever possible, as well as new instruments for which such observations are more frequently feasible.

This study also lays the groundwork for a broad range of further theoretical exploration. As detailed in Section 6, future efforts can target the M87* magnetosphere by explicitly treating synchrotron radiative cooling; firmer connections to gamma-ray binaries can be made by considering global properties of irradiated striped pulsar winds. Additional modelling may also verify the potential existence, outlined here (Section 5), of a Klein–Nishina reconnection regime featuring copious electron–positron pair production, which would have important implications for systems with a high fiducial cold magnetization, $\sigma_{\rm c,0} \gg \gamma_{\rm pp,\,max}/4$. However, even with a more modest order-unity pair yield (realized, in our simulations, when $\gamma_{\rm pp,\,max}/4\sigma_{\rm c,0} \lesssim 0.25$), Klein–Nishina reconnection could still function as an important source of *in situ* antimatter, taking, for example, an initially electron–proton plasma and injecting a prominent positron count. Such potential modifications to the composition of the reconnection plasma environment can be rigorously probed by applying the Klein–Nishina radiative physics studied here to the case of an initial electron–ion plasma.

More broadly, the theoretical framework established by this work (and the preceding article, Mehlhaff et al. 2021), especially the systematic recasting of Klein–Nishina physics as dimensionless energy scales, provides a paradigm for coupling the same physics to other particle acceleration processes (e.g. turbulence and shocks). That is, apart from factors of the reconnection rate, $\beta_{\rm rec} = 0.1$, entering into $\gamma_{\rm max}$ and $\gamma_{\rm rad,\,T}$, the Klein–Nishina scale hierarchy (50) also applies to other plasma physical processes, indicating when they, too, may couple to Klein–Nishina effects. The physical and observable consequences of this coupling, however, may be very different than in the context of reconnection as uncovered here and can only be elucidated via future dedicated simulations. This work thus represents only the first step toward a much larger suite of numerical investigations into Klein–Nishina-coupled particle acceleration, and, even more broadly, constitutes one of a growing number of theoretical and technological stepping stones (e.g. Hakobyan et al. 2019, 2023b; Schoeffler et al. 2019, 2023; Crinquand et al. 2020, 2021; Cruz et al. 2021; Crinquand et al. 2022; Chen et al. 2023; Groselj et al. 2023; Xie et al. 2023) toward a richer understanding of QED-coupled plasma physics in high-energy astrophysical environments.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Guillaume Dubus, Gilles Henri, and Hayk Hakobyan for stimulating discussions. The authors also extend their gratitude to the anonymous referee for a thorough review that greatly improved the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 863412). This work was also supported by NASA and the NSF, grant nos NASA ATP NNX17AK57G, NASA ATP 80NSSC20K0545, NASA ATP 80NSSC22K0828, NSF AST-1806084, and NSF AST-1903335. This research has made use of the TeVCat online source catalog (http://tevcat.uchicago.edu).

DATA AVAILABILITY

The simulation data underlying this article were generated at the XSEDE/TACC STAMPEDE2 supercomputer and are archived at the TACC/Ranch storage facility. As long as the data remain in the archive, they will be shared upon reasonable request to the corresponding author. The code used to generate the data, ZELTRON, is available at https://ipag.osug.fr/~ceruttbe/Zeltron/. The public version of the code does not yet contain the QED module underlying this work; a QED-enabled version will be shared upon reasonable request to the corresponding author.

REFERENCES

Abdo A. A. et al., 2009a, Science, 325, 840
Abdo A. A. et al., 2009b, ApJ, 701, L123
Abdo A. A. et al., 2009c, ApJ, 706, L56
Abdo A. A. et al., 2011, ApJ, 736, L11
Abdollahi S. et al., 2020, ApJS, 247, 33
Abdollahi S. et al., 2023, ApJS, 265, 31
Abeysekara A. U. et al., 2015, ApJ, 815, L22
Abramowski A. et al., 2012, ApJ, 746, 151
Acciari V. A. et al., 2009, Science, 325, 444
Acciari V. A. et al., 2010, ApJ, 716, 819
Ackermann M. et al., 2015, ApJ, 810, 14
Ackermann M. et al., 2016, ApJ, 824, L20
Aharonian F. et al., 2006, Science, 314, 1424
Aharonian F. et al., 2007, ApJ, 664, L71
Aharonian F. et al., 2009, A&A, 502, 749
Ahnen M. L. et al., 2016, A&A, 595, A98
Ahnen M. L. et al., 2017, A&A, 603, A29
Ajello M. et al., 2020, ApJ, 892, 105
Aleksić J. et al., 2011, ApJ, 730, L8
Aleksić J. et al., 2014, A&A, 569, A46
Aliu E. et al., 2012, ApJ, 746, 141
Atwood W. B. et al., 2009, ApJ, 697, 1071
Avara M. J., McKinney J. C., Reynolds C. S., 2016, MNRAS, 462, 636
Ball L., Dodd J., 2001, Publ. Astron. Soc. Aust., 18, 98
Ball L., Kirk J. G., 2000, Astropart. Phys., 12, 335
Ball D., Özel F., Psaltis D., Chan C.-k., 2016, ApJ, 826, 77
Ball D., Sironi L., Özel F., 2018, ApJ, 862, 80
Begelman M. C., Sikora M., 1987, ApJ, 322, 650
Beilicke M., VERITAS Collaboration, 2012, in Aharonian F. A., Hofmann W., Rieger F. M., eds, AIP Conf. Ser. Vol. 1505, High Energy Gamma-Ray Astronomy: 5th International Meeting on High Energy Gamma-Ray Astronomy. Am. Inst. Phys., New York, p. 586
Beloborodov A. M., 2017, ApJ, 850, 141
Bird G. A., 1989, Prog. Astronaut. Aeronaut., 117, 211
Birdsall C. K., 1991, IEEE Trans. Plasma Sci., 19, 65
Birdsall C. K., Langdon A. B., 1991, Plasma Physics via Computer Simulation. Adam Hilger, Bristol, England
Blackman E. G., Field G. B., 1994, Phys. Rev. Lett., 72, 494
Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467
Bloom S. D., Marscher A. P., 1996, ApJ, 461, 657
Blumenthal G. R., Gould R. J., 1970, Rev. Mod. Phys., 42, 237
Boeuf J. P., Marode E., 1982, J. Phys. D Appl. Phys., 15, 2169
Bogovalov S. V., 1999, A&A, 349, 1017
Caliandro G. A., Cheung C. C., Li J., Scargle J. D., Torres D. F., Wood K. S., Chernyakova M., 2015, ApJ, 811, 68
Cerutti B., Philippov A. A., 2017, A&A, 607, A134
Cerutti B., Werner G., 2019, Astrophysics Source Code Library, record ascl:1911.012
Cerutti B., Dubus G., Henri G., 2008, A&A, 488, 37
Cerutti B., Uzdensky D. A., Begelman M. C., 2012a, ApJ, 746, 148
Cerutti B., Werner G. R., Uzdensky D. A., Begelman M. C., 2012b, ApJ, 754, L33

Cerutti B., Werner G. R., Uzdensky D. A., Begelman M. C., 2013, ApJ, 770, 147
Cerutti B., Werner G. R., Uzdensky D. A., Begelman M. C., 2014a, Phys. Plasmas, 21, 056501
Cerutti B., Werner G. R., Uzdensky D. A., Begelman M. C., 2014b, ApJ, 782, 104
Cerutti B., Philippov A. A., Dubus G., 2020, A&A, 642, A204
Chang Z., Zhang S., Chen Y.-P., Ji L., Kong L.-D., Wang P.-J., 2021, Universe, 7, 472
Chashkina A., Bromberg O., Levinson A., 2021, MNRAS, 508, 1241
Chen A. Y., Uzdensky D., Dexter J., 2023, ApJ, 944, 173
Cherenkov Telescope Array Consortium et al., 2019, Science with the Cherenkov Telescope Array. World Scientific Publishing Co. Pte. Ltd., USA
Chernoglazov A., Hakobyan H., Philippov A. A., 2023, ApJ, 959, 122
Chernyakova M., Malyshev D., 2020, in Multifrequency Behaviour of High Energy Cosmic Sources – XIII. 3–8 June 2019. Palermo. p. 45 , preprint (arXiv:2006.03615)
Chernyakova M. et al., 2019, A&A, 631, A177
Chow A., Rowan M. E., Sironi L., Davelaar J., Bodo G., Narayan R., 2023a, MNRAS, 524, 90
Chow A., Davelaar J., Rowan M. E., Sironi L., 2023b, ApJ, 951, L23
Christie I. M., Petropoulou M., Sironi L., Giannios D., 2019, MNRAS, 482, 65
Christie I. M., Petropoulou M., Sironi L., Giannios D., 2020, MNRAS, 492, 549
Comisso L., Sironi L., 2021, Phys. Rev. Lett., 127, 255102
Coroniti F. V., 1990, ApJ, 349, 538
Costamante L., Cutini S., Tosti G., Antolini E., Tramacere A., 2018, MNRAS, 477, 4749
Crinquand B., Cerutti B., Philippov A., Parfrey K., Dubus G., 2020, Phys. Rev. Lett., 124, 145101
Crinquand B., Cerutti B., Dubus G., Parfrey K., Philippov A., 2021, A&A, 650, A163
Crinquand B., Cerutti B., Dubus G., Parfrey K., Philippov A., 2022, Phys. Rev. Lett., 129, 205101
Cruz F., Grismayer T., Chen A. Y., Spitkovsky A., Silva L. O., 2021, ApJ, 919, L4
Del Gaudio F., Grismayer T., Fonseca R. A., Silva L. O., 2020, J. Plasma Phys., 86, 905860516
Derouillat J. et al., 2018, Comput. Phys. Commun., 222, 351
Di Matteo T., 1998, MNRAS, 299, L15
Dubus G., 2006, A&A, 456, 801
Dubus G., 2013, A&AR, 21, 64
Dubus G., Guillard N., Petrucci P.-O., Martin P., 2017, A&A, 608, A59
EHT MWL Science Working Group et al., 2021, ApJ, 911, L11
El Mellah I., Cerutti B., Crinquand B., Parfrey K., 2022, A&A, 663, A169
El Mellah I., Cerutti B., Crinquand B., 2023, A&A, 677, A67
Esirkepov T. Z., 2001, Comput. Phys. Commun., 135, 144
Event Horizon Telescope Collaboration, 2019, ApJ, 875, L6
Event Horizon Telescope Collaboration, 2021, ApJ, 910, L13
Fedeli L. et al., 2022, in SC22: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society, Los Alamitos, CA, USA, p. 1
Fermi LAT Collaboration, 2012, Science, 335, 189
Fonseca R. A. et al., 2002, in Sloot P. M. A., Hoekstra A. G., Tan C. J. K., Dongarra J. J., eds, Computational Science – ICCS 2002. Springer, Berlin, Heidelberg, p. 342
Fossati G., Maraschi L., Celotti A., Comastri A., Ghisellini G., 1998, MNRAS, 299, 433
French O., Guo F., Zhang Q., Uzdensky D. A., 2023, ApJ, 948, 19
Galeev A. A., Rosner R., Vaiana G. S., 1979, ApJ, 229, 318
Galishnikova A., Philippov A., Quataert E., Bacchini F., Parfrey K., Ripperda B., 2023, Phys. Rev. Lett., 130, 115201
Gebhardt K., Adams J., Richstone D., Lauer T. R., Faber S. M., Gültekin K., Murphy J., Tremaine S., 2011, ApJ, 729, 119
Ghisellini G., 2011, in Aharonian F. A., Hofmann W., Rieger F. M., eds, AIP Conf. Proc. Vol. 1381, 25th Texas Symposium on Relativistic AstroPhysics (Texas 2010). Am. Inst. Phys., New York, p. 180
Ghisellini G., Righi C., Costamante L., Tavecchio F., 2017, MNRAS, 469, 255
Giannios D., 2013, MNRAS, 431, 355
Giannios D., Uzdensky D. A., 2019, MNRAS, 484, 1378
Giannios D., Uzdensky D. A., Begelman M. C., 2009, MNRAS, 395, L29
Giannios D., Uzdensky D. A., Begelman M. C., 2010, MNRAS, 402, 1649
Gierliński M., Zdziarski A. A., 2003, MNRAS, 343, L84
Goodbred M., Liu Y.-H., 2022, Phys. Rev. Lett., 129, 265101
Goodman J., Uzdensky D., 2008, ApJ, 688, 555
Gould R. J., Schréder G. P., 1967, Phys. Rev., 155, 1404
Greenwood A. D., Cartwright K. L., Luginsland J. W., Baca E. A., 2004, J. Comput. Phys., 201, 665
Grismayer T., Vranic M., Martins J. L., Fonseca R. A., Silva L. O., 2016, Phys. Plasmas, 23, 056706
Groselj D., Hakobyan H., Beloborodov A. M., Sironi L., Philippov A., 2023, preprint (arXiv:2301.11327)
Guo F., Li H., Daughton W., Liu Y.-H., 2014, Phys. Rev. Lett., 113, 155005
Guo F., Liu Y.-H., Daughton W., Li H., 2015, ApJ, 806, 167
Guo F., Li X., Daughton W., Kilian P., Li H., Liu Y.-H., Yan W., Ma D., 2019, ApJ, 879, L23
Guo F., Li X., Daughton W., Li H., Kilian P., Liu Y.-H., Zhang Q., Zhang H., 2021, ApJ, 919, 111
H. E. S. S. Collaboration, 2019, A&A, 627, A159
H. E. S. S. Collaboration, 2020, A&A, 633, A162
H. E. S. S. Collaboration, 2021, A&A, 648, A23
Hakobyan H., Philippov A., Spitkovsky A., 2019, ApJ, 877, 53
Hakobyan H., Petropoulou M., Spitkovsky A., Sironi L., 2021, ApJ, 912, 48
Hakobyan H., Spitkovsky A., Chernoglazov A., Philippov A., Groselj D., Mahlmann J., 2023a, PrincetonUniversity/tristan-mp-v2: v2.6, Zenodo. Available at:https://zenodo.org/records/7566725.
Hakobyan H., Ripperda B., Philippov A. A., 2023b, ApJ, 943, L29
Harrison M. G., Neukirch T., 2009, Phys. Rev. Lett., 102, 135003
Hayashida M. et al., 2012, ApJ, 754, 114
Hayashida M. et al., 2015, ApJ, 807, 79
Hockney R. W., Eastwood J. W., 1988, Computer Simulation using Particles. Hilger, Bristol
Hoshino M., Lyubarsky Y., 2012, Space Sci. Rev., 173, 521
Igumenshchev I. V., 2008, ApJ, 677, 317
Johnston S., Manchester R. N., Lyne A. G., Bailes M., Kaspi V. M., Qiao G., D'Amico N., 1992, ApJ, 387, L37
Jones F. C., 1968, Phys. Rev., 167, 1159
Khangulyan D., Aharonian F. A., Bogovalov S. V., Ribó M., 2012, ApJ, 752, L17
Kim J. Y. et al., 2018, A&A, 616, A188
Kimura S. S., Toma K., Noda H., Hada K., 2022, ApJ, 937, L34
Kirk J. G., Skjæraasen O., 2003, ApJ, 591, 366
Kirk J. G., Ball L., Skjæraasen O., 1999, Astropart. Phys., 10, 31
Kirk J. G., Lyubarsky Y., Petri J., 2009, in Becker W.ed., Astrophysics and Space Science Library, Vol. 357, Neutron Stars and Pulsars. Springer, Berlin Heidelberg, p. 421
Levinson A., Cerutti B., 2018, A&A, 616, A184
Li Y.-P., Yuan F., Wang Q. D., 2017, MNRAS, 468, 2552
Li X., Guo F., Li H., Li S., 2018, ApJ, 866, 4
Lin S. L., Bardsley J. N., 1977, J. Chem. Phys., 66, 435
Ly C., Walker R. C., Junor W., 2007, ApJ, 660, 200
Lyubarsky Y. E., 2003, MNRAS, 345, 153
Lyubarsky Y. E., 2005, MNRAS, 358, 113
Lyubarsky Y., Kirk J. G., 2001, ApJ, 547, 437
Lyutikov M., Uzdensky D., 2003, ApJ, 589, 893
MAGIC Collaboration, 2008, Science, 320, 1752
MAGIC Collaboration, 2018, A&A, 619, A159
MAGIC Collaboration, 2020, MNRAS, 492, 5354

MAGIC Collaboration, 2021, A&A, 647, A163
Madejski G. G., Sikora M., 2016, ARA&A, 54, 725
Maraschi L., Ghisellini G., Celotti A., 1992, ApJ, 397, L5
Mehlhaff J. M., Werner G. R., Uzdensky D. A., Begelman M. C., 2020, MNRAS, 498, 799
Mehlhaff J. M., Werner G. R., Uzdensky D. A., Begelman M. C., 2021, MNRAS, 508, 4532
Melia F., Königl A., 1989, ApJ, 340, 162
Melzani M., Walder R., Folini D., Winisdoerffer C., Favre J. M., 2014, A&A, 570, A112
Meyer M., Scargle J. D., Blandford R. D., 2019, ApJ, 877, 39
Michel F. C., 1994, ApJ, 431, 397
Mizuno Y., 2022, Universe, 8, 85
Moderski R., Sikora M., Coppi P. S., Aharonian F., 2005, MNRAS, 363, 954
Nalewajko K., 2016, Galaxies, 4, 28
Nalewajko K., Giannios D., Begelman M. C., Uzdensky D. A., Sikora M., 2011, MNRAS, 413, 333
Nalewajko K., Begelman M. C., Cerutti B., Uzdensky D. A., Sikora M., 2012, MNRAS, 425, 2519
Nalewajko K., Yuan Y., Chruślińska M., 2018, J. Plasma Phys., 84, 755840301
Nättilä J., Beloborodov A. M., 2021, ApJ, 921, 87
Nenkova M., Sirocky M. M., Ivezić Ž., Elitzur M., 2008a, ApJ, 685, 147
Nenkova M., Sirocky M. M., Nikutta R., Ivezić Ž., Elitzur M., 2008b, ApJ, 685, 160
Neukirch T., Wilson F., Harrison M. G., 2009, Phys. Plasmas, 16, 122102
Olguín-Iglesias A. et al., 2016, MNRAS, 460, 3202
Ortuño-Macías J., Nalewajko K., 2020, MNRAS, 497, 1365
Petropoulou M., Sironi L., 2018, MNRAS, 481, 5687
Petropoulou M., Giannios D., Sironi L., 2016, MNRAS, 462, 3325
Philippov A. A., Spitkovsky A., 2018, ApJ, 855, 94
Porth O., Mizuno Y., Younsi Z., Fromm C. M., 2021, MNRAS, 502, 2023
Pozdnyakov L. A., Sobol I. M., Syunyaev R. A., 1983, Astrophys. Space Phys. Res., 2, 189
Prandini E., Ghisellini G., 2022, Galaxies, 10, 35
Rees H. D., 1968, Phys. Lett. A, 26, 416
Ripperda B., Bacchini F., Philippov A. A., 2020, ApJ, 900, 100
Ripperda B., Liska M., Chatterjee K., Musoke G., Philippov A. A., Markoff S. B., Tchekhovskoy A., Younsi Z., 2022, ApJ, 924, L32
Rybicki G. B., Lightman A. P., 1979, Radiative Processes in Astrophysics. Wiley, New York
Scepi N., Dexter J., Begelman M. C., 2022, MNRAS, 511, 3536
Schoeffler K. M., Grismayer T., Uzdensky D., Fonseca R. A., Silva L. O., 2019, ApJ, 870, 49
Schoeffler K. M., Grismayer T., Uzdensky D., Silva L. O., 2023, MNRAS, 523, 3812
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Sierpowska A., Bednarek W., 2005, MNRAS, 356, 711
Sierpowska-Bartosik A., Bednarek W., 2008, MNRAS, 385, 2279
Sikora M., Begelman M. C., Rees M. J., 1994, ApJ, 421, 153
Sikora M., Stawarz Ł., Moderski R., Nalewajko K., Madejski G. M., 2009, ApJ, 704, 38
Sironi L., 2022, Phys. Rev. Lett., 128, 145102
Sironi L., Beloborodov A. M., 2020, ApJ, 899, 52
Sironi L., Spitkovsky A., 2014, ApJ, 783, L21
Sironi L., Giannios D., Petropoulou M., 2016, MNRAS, 462, 48
Sironi L., Rowan M. E., Narayan R., 2021, ApJ, 907, L44
Sitarek J. et al., 2015, in 34th International Cosmic Ray Conference (ICRC2015). p. 825 , preprint (arXiv:1508.04580)
Sobacchi E., Nättilä J., Sironi L., 2021, MNRAS, 503, 688
Sridhar N., Sironi L., Beloborodov A. M., 2021, MNRAS, 507, 5625
Sridhar N., Sironi L., Beloborodov A. M., 2023, MNRAS, 518, 1301
Tam P. H. T., Huang R. H. H., Takata J., Hui C. Y., Kong A. K. H., Cheng K. S., 2011, ApJ, 736, L10
Tam P. H. T., He X. B., Pal P. S., Cui Y., 2018, ApJ, 862, 165
Tamburini M., Pegoraro F., Di Piazza A., Keitel C. H., Macchi A., 2010, New J. Phys., 12, 123005
Tavecchio F., Ghisellini G., 2008, MNRAS, 386, 945
Tchekhovskoy A., Narayan R., McKinney J. C., 2011, MNRAS, 418, L79
Timokhin A. N., Harding A. K., 2015, ApJ, 810, 144
Timokhin A. N., Harding A. K., 2019, ApJ, 871, 12
Uzdensky D. A., 2011, Space Sci. Rev., 160, 45
Uzdensky D. A., 2016, Radiative Magnetic Reconnection in Astrophysics, Vol. 427. Springer International Publishing, Cham, p. 473
Uzdensky D. A., 2022, J. Plasma Phys., 88, 905880114
Uzdensky D. A., Goodman J., 2008, ApJ, 682, 608
Uzdensky D. A., Loureiro N. F., Schekochihin A. A., 2010, Phys. Rev. Lett., 105, 235002
Uzdensky D. et al., 2019, BAAS, 51, 362
Vranic M., Grismayer T., Martins J. L., Fonseca R. A., Silva L. O., 2015, Comput. Phys. Commun., 191, 65
Wakely S. P., Horan D., 2008, in Caballero R., D'Olivo J. C., Medina-Tanco G., Nellen L., Sánchez F. A., Valdés-Galicia J. F., eds, International Cosmic Ray Conference, Vol. 3. Universidad Nacional Autónoma de México, Mexico City, Mexico, p. 1341, Available at: http://tevcat.uchicago.edu
Walker R. C., Hardee P. E., Davies F. B., Ly C., Junor W., 2018, ApJ, 855, 128
Watanabe N., Yokoyama T., 2006, ApJ, 647, L123
Werner G. R., Uzdensky D. A., 2017, ApJ, 843, L27
Werner G. R., Uzdensky D. A., Cerutti B., Nalewajko K., Begelman M. C., 2016, ApJ, 816, L8
Werner G. R., Uzdensky D. A., Begelman M. C., Cerutti B., Nalewajko K., 2018, MNRAS, 473, 4840
Werner G. R., Philippov A. A., Uzdensky D. A., 2019, MNRAS, 482, L60
Xie Y. et al., 2023, Sci. Bull., 68, 1857
Yao P. Z., Dexter J., Chen A. Y., Ryan B. R., Wong G. N., 2021, MNRAS, 507, 4864
Zacharias M. et al., 2017, in AIP Conf. Proc. Vol. 1792, 6th International Symposium on High Energy Gamma-Ray Astronomy. Am. Inst. Phys., New York, p. 050023
Zdziarski A. A., Malyshev D., Chernyakova M., Pooley G. G., 2017, MNRAS, 471, 3657
Zenitani S., Hoshino M., 2001, ApJ, 562, L63
Zenitani S., Hoshino M., 2008, ApJ, 677, 530
Zhang H., Sironi L., Giannios D., 2021, ApJ, 922, 261
Zhang H., Li X., Giannios D., Guo F., Thiersen H., Böttcher M., Lewis T., Venters T., 2022, ApJ, 924, 90
Zhang H., Sironi L., Giannios D., Petropoulou M., 2023, ApJ, 956, L36
Zhdankin V., Uzdensky D. A., Werner G. R., Begelman M. C., 2020, MNRAS, 493, 603
Zhdankin V., Uzdensky D. A., Kunz M. W., 2021, ApJ, 908, 71

APPENDIX A: APPROXIMATE SCATTERED PHOTON ENERGY ACCURACY

Here, we demonstrate that approximation (26) preserves the average energy, $\langle \epsilon_{\rm scat} \rangle$, to which seed photons are IC scattered by a particle of a given Lorentz factor, γ. This is done in Fig. A1, where we compare the exact mean scattered photon energy to the one obtained after many Monte Carlo trials using equation (26). The curves overlap well across the presented range of $\gamma/\gamma_{\rm KN}$, which covers the entire range of $\gamma/\gamma_{\rm KN}$ encountered across all simulations presented in the main text. Because the total cross-section (23) is not approximated, the preservation of $\langle \epsilon_{\rm scat} \rangle$ demonstrated here also implies the correct total IC power (42).

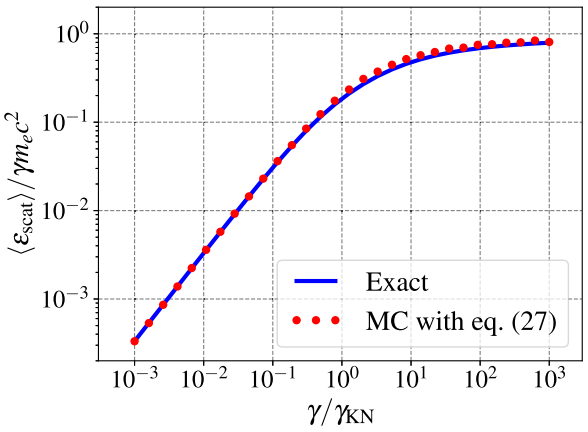

Figure A1. The exact (solid line) and approximate (points) mean final photon energy, $\langle \epsilon_{\rm scat} \rangle$, after a seed photon (energy $\epsilon_{\rm ph}$) undergoes IC scattering with a particle of Lorentz factor γ. The exact curve (cf. fig. 3 of Mehlhaff et al. 2021) is computed by dividing equation (42) by equation (5). The approximate data are obtained by averaging over 10^5 Monte Carlo trials per γ value, each one using the procedure described in Section 2.2.3 that employs approximation (26).

APPENDIX B: LIKELIHOOD OF LIMIT CYCLES

Here, we expand on Section 4.6, using measurements from our IC(KN)+PP run (Section 3) to fill in the main quantitative uncertainties from the work of Mehlhaff et al. (2021) on the pair-loading efficiency – the parameter ξ in what follows – including its implications for the pair-loaded magnetization, $\sigma_{\rm h}^*$, and on the possibility of $\sigma_{\rm h}$-mediated limit cycles. We report first the expressions for the produced-particle energy density, $U_{\rm prod.}$, and the pair-loaded hot magnetization, $\sigma_{\rm h}$, flowing into the reconnection layer from the upstream region (thus, in the context of Fig. 14, both quantities should be evaluated along, or perhaps just upstream of, the reconnection separatrix). Mehlhaff et al. (2021) found that $U_{\rm prod.}$ evaluated at this location can be written as

$$U_{\rm prod.} = \xi \mathcal{F} \frac{B_0^2}{8\pi} \,. \tag{B1}$$

Here, $\mathcal{F}$ is the fraction of the IC power, emitted by particles in the reconnection layer, that is radiated above pair-production threshold, and ξ is the pair-loading efficiency, which we refer to from here onward according to a more precise name signifying its role in equation (B1), the *energy recapture efficiency*: how much of the above-threshold radiation is recaptured by the reconnection layer from the inflow region as hot pairs. Accounting for the newborn pair energy density $U_{\rm prod.}$, the hot magnetization, $\sigma_{\rm h}$, can be written in terms of $\mathcal{F}$ and ξ as

$$\sigma_{\rm h} = \frac{\sigma_{\rm h,0}}{1 + 2\mathcal{F}\xi\sigma_{\rm h,0}/3} \simeq \frac{3}{2\mathcal{F}\xi} \,, \tag{B2}$$

where the approximation holds when $\mathcal{F}\xi\sigma_{\rm h,0} \gg 1$. Provided that the reconnection layer responds only to the present – as opposed to the past – magnetization of the plasma feeding it, equation (B2) encodes a universal ($\sigma_{\rm h,0}$-independent) *pair-loaded magnetization*,

$$\sigma_{\rm h}^* \equiv \frac{3}{2\mathcal{F}\xi} \,, \tag{B3}$$

which we first introduced qualitatively in Section 4.6. The pair-loaded magnetization $\sigma_{\rm h}^*$ is determined self-consistently by the (potentially $\sigma_{\rm h}^*$-dependent) values of $\mathcal{F}$ and ξ. As discussed in Section 4.6, any Klein–Nishina reconnection layer with initial magnetization $\sigma_{\rm h,0} > \sigma_{\rm h}^*$ will try to self-regulate to $\sigma_{\rm h}^*$, but, if the self-regulation is too efficient, the system will enter a limit cycle between a high and low magnetization on either side of $\sigma_{\rm h}^*$.

One major uncertainty in the model developed by Mehlhaff et al. (2021) was the energy recapture efficiency, ξ. This parameter is particularly important for two reasons. First, it is one of the key quantities deciding the pair-loaded magnetization, $\sigma_{\rm h}^*$. Second, it determines how strongly the reconnection layer is coupled, via the upstream plasma, to its own NTPA, and, in particular, whether this coupling is strong enough to drive the system into a limit cycle. For a wide range of assumptions, Mehlhaff et al. (2021) found that limit cycles require ξ of order unity – at least 0.3 or so in the most lenient case, but often even higher. In order to make contact with these two issues – the value of $\sigma_{\rm h}^*$ and the existence (or not) of limit cycles – we now proceed to measure ξ and $\sigma_{\rm h}^*$ from our simulation.

We estimate the above-threshold power fraction, $\mathcal{F}$, as the part of the average IC emission spectrum in Fig. 9 radiated above pair-production threshold, measuring $\mathcal{F} \simeq 0.5$. This agrees with the prediction one arrives at by taking, from Fig. 7, a particle distribution power law, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-2}$, with a cut-off at $200\gamma_{\rm KN}$ and consulting the corresponding $\mathcal{F}$-value from fig. 10 of Mehlhaff et al. (2021). Armed with $\mathcal{F}$, we can estimate ξ by measuring $U_{\rm prod.}$ along the reconnection separatrix in Fig. 14 and plugging the result into equation (B1). From the figure, we see that, once the pair coat is built up around the reconnection layer, it presents an inflowing produced-particle energy density of about $0.01 B_0^2/8\pi$ at the separatrix crossing. This implies that $\xi\mathcal{F} \sim 0.01$ and, hence, $\xi \sim 0.02$. Finally, inserting $\xi\mathcal{F} \sim 0.01$ into equation (B3) yields $\sigma_{\rm h}^* \sim 150$. This is about a factor of 10 smaller than $\sigma_{\rm h,0} = 1250$ in the IC(KN)+PP run, and, hence, the contour $\sigma_{\rm h} = 0.1\sigma_{\rm h,0}$ in Fig. 14 nearly overlaps the $U_{\rm prod.} = 0.01 B_0^2/8\pi$ contour.[8]

According to Mehlhaff et al. (2021), our measured value of ξ is much too low (by about an order of magnitude) for limit cycles to occur: the coupling between the inflow region and the layer is too mild. This is consistent with our simulations, from which we identify no evidence of cyclic behaviour, neither in the reconnection-powered NTPA, nor in the pair loading and resulting magnetization, $\sigma_{\rm h}$, presented to the reconnection layer.

Let us now examine how our measured value of ξ, including its implications on the existence of limit cycles, may generalize under changes of the reconnection parameters (such as to those of real astrophysical systems). To inform this discussion, we summarize here the basic physics that determines ξ. Mehlhaff et al. (2021) explain that ξ can be written as the product, $\xi = f_{\rm inj} f_{\rm nocool} f_{\rm noesc}$, where the three factors on the right-hand side correspond to the three main loss mechanisms that inhibit above-threshold radiation from being recaptured by the reconnection layer from the upstream region. First, $f_{\rm inj}$ is the fraction of the layer-produced above-threshold radiation that successfully traverses the reconnection separatrix into the upstream region. The factor $f_{\rm inj}$ falls below unity when some photons – for example, in the case of extreme beaming of reconnection-accelerated particles along the reconnection layer (e.g. Cerutti, Uzdensky & Begelman 2012a; Cerutti et al. 2012b, 2013, 2014a, b; Mehlhaff

[8]In fact, the $\sigma_{\rm h} = 0.1\sigma_{\rm h,0}$ contour lies slightly farther from the main reconnection X-point than the one for $U_{\rm prod.} = 0.01 B_0^2/8\pi$. The small discrepancy originates from the in-plane magnetic field slightly weakening near reconnection X-points, an effect visible in the cold magnetization maps of Fig. 14. This effect owes to the inflowing magnetic field lines draping themselves across plasmoids, hanging from them like the cables of a suspension bridge, and, hence, thinning out as they sink toward the X-points in between.

et al. 2020) – produce pairs in the downstream region.[9] Second, $f_{\rm nocool}$ is that part of the energy deposited into the upstream region as newborn pairs that is not radiated away while those pairs are re-advected toward the layer. Finally, $f_{\rm noesc}$ is the fraction of newborn upstream particles that do not escape the system (e.g. by travelling along a field line in the $\pm x$-directions) before being swept (in the $\pm y$-directions) into the layer.

To determine how ξ may change with the reconnection parameters, we sketch the dependence of the three governing factors $f_{\rm inj}$, $f_{\rm nocool}$, and $f_{\rm noesc}$, on these parameters as revealed both by analytic theory (Mehlhaff et al. 2021) and by our simulations. Given the periodic boundaries of our setup, particle escape is impossible. Hence, $f_{\rm noesc}$, perhaps artificially, equals 1 in the simulations, which are then restricted to probing $f_{\rm nocool}$ and $f_{\rm inj}$. Of these two, Mehlhaff et al. (2021) provide analytic estimates of $f_{\rm nocool}$, showing that, similarly to $\mathcal{F}$, it depends only on the shape of the reconnection-energized particle distribution – for a power law, $\mathrm{d}N/\mathrm{d}\gamma \propto \gamma^{-p}$, on the index, p, and on the high-energy cut-off, $\gamma_{\rm c}$, normalized by $\gamma_{\rm KN}$. Those authors found that, for $p > 2$, $f_{\rm nocool}$ is controlled by the low-energy particles, and its value, $f_{\rm nocool} \sim 0.1$, is therefore independent of $\gamma_{\rm c}$. For $p < 2$, $f_{\rm nocool}$ acquires a weak dependence on $\gamma_{\rm c}$ – signalling the enhanced importance of particles in the high-energy tail – but even then does not reach order unity unless $\gamma_{\rm c} \gg 10^3\gamma_{\rm KN}$ and $p \simeq 1$. In the context of our IC(KN)+PP run, the Mehlhaff et al. (2021) estimates (e.g. their fig. B2) suggest that $f_{\rm nocool} \sim 0.1$ for $p \simeq 2$ and $\gamma_{\rm c} \sim 200\gamma_{\rm KN}$ (as in Fig. 7). If we adopt this value, we can derive the empirical measurement, $f_{\rm inj} \sim \xi/f_{\rm nocool} \sim 0.2$.

Now, while $f_{\rm nocool}$ depends on the details of reconnection-powered NTPA, $f_{\rm noesc}$ and $f_{\rm inj}$ are dictated instead mostly by the large-scale geometry and kinematics of reconnection. For example, $f_{\rm noesc}$ depends on the free-streaming time, $\lesssim L/c$, for a newborn upstream particle to vacate the system by following an unreconnected magnetic field line (a more complete discussion of the factors influencing $f_{\rm noesc}$ can be found in the appendix C of Mehlhaff et al. 2021). As another example, the factor, $f_{\rm inj}$, depends on the effective width of the reconnection layer and on beaming, both that associated with bulk (e.g. plasmoid-chain) motion and that stemming from kinetic effects near reconnection X-points. These remarks suggest that $f_{\rm noesc}$ and $f_{\rm inj}$ are constants (at least for $\sigma_{\rm h,0} \gg 1$), because the processes deciding them are either generic byproducts of relativistic reconnection (in the case of kinetic beaming; Cerutti et al. 2012b; Mehlhaff et al. 2020) or tied to the large-scale evolution (e.g. of the plasmoid chain), which seems rather insensitive to Klein–Nishina radiative physics, as discussed in Sections 4.1 and 4.2.

This, then, makes clear the utility of introducing the individual factors $f_{\rm inj}$, $f_{\rm nocool}$, and $f_{\rm noesc}$: two of them, $f_{\rm noesc}$ and $f_{\rm inj}$, are expected to be roughly constant regardless of the exact reconnection parameter values (as long as we are in the Klein–Nishina reconnection regime), and the remaining factor, $f_{\rm nocool}$, is one that we can estimate based on a substantially reduced set of parameters – the NTPA power-law index, p, and cut-off energy, $\gamma_{\rm c}$. This enables us, as we seek to extrapolate our simulation results to reason about ξ and $\sigma_{\rm h}^*$ in astrophysical Klein–Nishina reconnection, to skip an exhaustive exploration of the high-dimensional radiative reconnection parameter space (e.g. all orderings of the energy scales, $\sigma_{\rm c}$, $\gamma_{\rm max}$, $\gamma_{\rm rad,\,T}$, $\gamma_{\rm cool}$, $\gamma_{\rm KN}$, etc.). Instead, we can focus on just p and $\gamma_{\rm c}$. While it is true that these two are still determined self-consistently by the much larger underlying parameter space, we leave a quantitative characterization of this dependence to future work, having already noted several general trends in Section 4.3.

The established logical framework in terms of $f_{\rm inj}$, $f_{\rm nocool}$, and $f_{\rm noesc}$ – with just the two independent variables p and $\gamma_{\rm c}$ – equips us to estimate a global upper bound on ξ. In the above-mentioned case of extreme NTPA wherein $p \simeq 1$, $\gamma_{\rm c} \gg 10^3\gamma_{\rm KN}$ and, thus, $f_{\rm nocool} \sim 1$, the combined factor, $\xi = f_{\rm nocool}f_{\rm inj}f_{\rm noesc}$, reduces to $\xi \sim f_{\rm inj}f_{\rm noesc} \leq f_{\rm inj} \sim 0.2$. Thus, even in the most favourable circumstances, ξ is still smaller than the most lenient minimum required value, 0.3, found by Mehlhaff et al. (2021) to produce limit cycles. We emphasize the critical role played by simulations in reaching this result, for they provide the necessary bound on the factor, $f_{\rm inj}$, that caps ξ to beneath the limit cycle value range. In view of this combined input from analytic theory and simulations, we speculate that *$\sigma_{\rm h}$-mediated limit cycles are unlikely, even in astrophysical instances of Klein–Nishina reconnection* .

Finally, let us comment on how varying system parameters may impact the pair-loaded magnetization, $\sigma_{\rm h}^* = 3/2\xi\mathcal{F}$. The discussion thus far can be translated into an expected lower bound on $\sigma_{\rm h}^*$ as follows. We have already seen that extreme NTPA ($p \simeq 1$ and $\gamma_{\rm c} \geq 10^3\gamma_{\rm KN}$) yields $f_{\rm nocool} \sim 1$ and, if one also favourably posits $f_{\rm noesc} = 1$, a global maximum, $\xi \sim 0.1$. In the same NTPA regime, the above-threshold power fraction, $\mathcal{F}$, attains order unity (fig. 10 of Mehlhaff et al. 2021). Thus, with $\xi\mathcal{F}$ as large as possible given numerical measurements and analytic expectations, we arrive at the minimum possible pair-loaded magnetization, $\sigma_{\rm h}^* \simeq 3/(2 \times 0.1 \times 1) \sim 15$.

If we now relax these extreme assumptions and allow for the (likely) possibility that NTPA is not quite as efficient, we may estimate a more typical value of $\sigma_{\rm h}^*$. As explained previously, once the power-law index of the particle distribution softens and/or the high-energy cut-off, $\gamma_{\rm c}$, falls below $\sim 10^3\gamma_{\rm KN}$, $f_{\rm nocool}$ loses dependence on these two parameters and becomes closer to 0.1. In the same regime (again referring to fig. 10 of Mehlhaff et al. 2021), $\mathcal{F}$ becomes more strongly dependent on NTPA, but still attains $\sim$0.5 when $p \leq 2$ for a broad range of $\gamma_{\rm c}$. This gives a more typical pair-loaded magnetization (realized in our simulation) of $\sigma_{\rm h}^* \sim 150(0.5/\mathcal{F})(1/f_{\rm noesc})$.

To summarize, our IC(KN)+PP run allows us to measure the all-important energy recapture efficiency, ξ, revealing a value that is too low, in the context of the work by Mehlhaff et al. (2021), to expect violent limit cycles mediated by pair regulation of $\sigma_{\rm h}$. Such limit cycles are consequently unlikely in real astrophysical systems. The expected evolutionary pathway taken by astrophysical Klein–Nishina reconnection (as long as the newborn pairs remain few in number) is then the same as that of our main IC(KN)+PP run: provided $\sigma_{\rm h,0} > \sigma_{\rm h}^*$, the system will self-regulate directly to the $\sigma_{\rm h,0}$-independent value, $\sigma_{\rm h}^*$, without overshooting. However, if $\sigma_{\rm h,0} < \sigma_{\rm h}^*$, the system remains at its initial magnetization.

[9]Mehlhaff et al. (2021) did not include $f_{\rm inj}$ explicitly in their model, but discuss its effect in their appendix C.

This paper has been typeset from a TEX/LATEX file prepared by the author.