Pair-regulated Klein–Nishina relativistic magnetic reconnection with applications to blazars and accreting black holes

J. M. Mehlhaff [1],[*] G. R. Werner,[1] D. A. Uzdensky[1] and M. C. Begelman[2,3]

[1]*Center for Integrated Plasma Studies, Physics Department, University of Colorado, 390 UCB, Boulder, CO 80309, USA*
[2]*JILA, University of Colorado and National Institute of Standards and Technology, 440 UCB, Boulder, CO 80309, USA*
[3]*Department of Astrophysical and Planetary Sciences, University of Colorado, 391 UCB, Boulder, CO 80309, USA*

ABSTRACT

Relativistic magnetic reconnection is a powerful agent through which magnetic energy can be tapped in astrophysics, energizing particles that then produce observed radiation. In some systems, the highest energy photons come from particles Comptonizing an ambient radiation bath supplied by an external source. If the emitting particle energies are high enough, this inverse Compton (IC) scattering enters the Klein–Nishina regime, which differs from the low-energy Thomson IC limit in two significant ways. First, radiative losses become inherently discrete, with particles delivering an order-unity fraction of their energies to single photons. Secondly, Comptonized photons may pair produce with the ambient radiation, opening up another channel for radiative feedback on magnetic reconnection. We analytically study externally illuminated highly magnetized reconnecting systems for which both of these effects are important. We identify a universal (initial magnetization-independent) quasi-steady state in which gamma-rays emitted from the reconnection layer are absorbed in the upstream region, and the resulting hot pairs dominate the energy density of the inflow plasma. However, a true pair cascade is unlikely, and the number density of created pairs remains subdominant to that of the original plasma for a wide parameter range. Future particle-in-cell simulation studies may test various aspects. Pair-regulated Klein–Nishina reconnection may explain steep spectra (quiescent and flaring) from flat-spectrum radio quasars and black hole accretion disc coronae.

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

1 INTRODUCTION

Many accreting and jet-launching compact objects host tenuous, highly magnetized plasmas that are prone to dissipation through collisionless relativistic magnetic reconnection (Blackman & Field 1994; Lyutikov & Uzdensky 2003; Lyubarsky 2005). Reconnection represents an important pathway that transfers free magnetic energy to plasma internal and kinetic energy, two forms that may then be channelled into observable radiation. In some circumstances, the two steps in this energy conversion process – from the magnetic field to plasma through reconnection, and from plasma to light through radiative processes – can be imagined as happening separately. This is true, for example, when the plasma radiative cooling time is longer than the characteristic time-scale on which reconnection occurs.

If, however, the emitting particles cool on time-scales comparable to – or even much shorter than – the reconnection dynamical time, then radiative and reconnection physics are inseparable. The coupling between them may, furthermore, be facilitated by much more than just optically thin radiative cooling, where the emitting particles suffer radiative drag but the produced photons passively escape the system. In some situations, the optical depths to various processes, including Thomson scattering and pair creation, may exceed unity, affording the emitted photons further opportunity to impact the ongoing reconnection process. Even when these optical depths are small – but especially when they are not – a self-consistent approach that models radiation and reconnection simultaneously is required to capture the modifications that the various radiative interactions may make to the reconnection-powered photon spectrum. In this paper, we term any regime of this kind, in which reconnection and photon processes are inextricably coupled, as a 'radiative' regime of magnetic reconnection (Uzdensky 2016).

Due to high magnetic and radiation energy densities, reconnection in relativistic compact object environments is likely to be highly radiative, and much recent work on radiative reconnection is related to these systems, including studies of pulsar winds (Pétri 2012; Cerutti & Philippov 2017; Cerutti, Philippov & Dubus 2020), pulsar wind nebulae (Uzdensky, Cerutti & Begelman 2011; Cerutti, Uzdensky & Begelman 2012a; Cerutti et al. 2013, 2014a, b; Yuan et al. 2016), pulsar magnetospheres (Lyubarskii 1996; Uzdensky & Spitkovsky 2014; Philippov, Spitkovsky & Cerutti 2015; Cerutti, Philippov & Spitkovsky 2016; Philippov & Spitkovsky 2018; Hakobyan, Philippov & Spitkovsky 2019), magnetar magnetospheres (Schoeffler et al. 2019), gamma-ray bursts (McKinney & Uzdensky 2012), accreting black holes (Beloborodov 2017; Werner, Philippov & Uzdensky 2019; Sironi & Beloborodov 2020), blazars (Nalewajko et al. 2011, 2012; Nalewajko, Yuan & Chruślińska 2018; Mehlhaff et al. 2020; Ortuño-Macías & Nalewajko 2020), and black hole magnetospheres (Parfrey, Philippov & Cerutti 2019; Crinquand et al. 2021). Some studies have not focused on a single object class, but have still been motivated by some combination of the above (e.g. Jaroschek & Hoshino 2009; Uzdensky 2011, 2016; Uzdensky
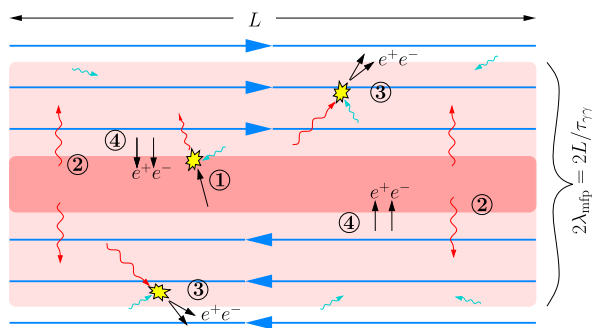
[*] E-mail: john.mehlhaff@colorado.edu

Figure 1. Schematic of Klein–Nishina radiative magnetic reconnection when the system is optically thick $\tau_{\gamma\gamma} \gg 1$ to pair production. In this regime, the mean free path λ_{mfp} of a high-energy photon is less than the system size L: $\lambda_{\mathrm{mfp}} = L/\tau_{\gamma\gamma} \ll L$. Thus, every high-energy photon (red wiggled arrows) annihilates against an ambient seed photon (cyan wiggled arrows) before escaping the system. Newborn pairs are denoted by black arrows and magnetic field lines by blue arrows. The reconnection layer is opaque pink; the region penetrated by high-energy photons (one mean free path away from the layer) is transparent pink. See the text for a description of stages 1–4.

& McKinney 2011; Nättilä & Beloborodov 2020; Hakobyan et al. 2021).

In many astrophysical contexts, the reconnection region is expected to be illuminated by an external source of soft photons – with energies much lower than the electron temperature – and inverse Compton (IC) scattering of these photons dominates the emitted light. For some such systems, observations further suggest that particles emitting at the highest photon energies do so in the Klein–Nishina regime. For example, TeV observations of the flat-spectrum radio quasar (FSRQ) PKS 1222+21 (Aleksić et al. 2011) indicate that the observed TeV photons, if Comptonized from radiation impinging on the jet from a hot dust region, are produced in the marginal Klein–Nishina regime of the IC process (Mehlhaff et al. 2020). Thus, if the radiating particles are accelerated by relativistic reconnection (for which a case has been made by Nalewajko et al. 2012; Mehlhaff et al. 2020), then it is likely that the collective reconnection dynamics are significantly impacted by Klein–Nishina IC effects. In an entirely different type of system, an X-ray binary, recent *Fermi* observations reveal that the spectral cut-off in the quiescent high-luminosity state of Cyg X-1 lies in the 40–80 MeV range (Zdziarski et al. 2017). As in the case of FSRQs, this hints that the most energetic particles Comptonize ambient photons (sourced, in this case, by the accretion disc) in the Klein–Nishina regime.

In addition to qualitatively modifying the radiative cooling experienced by particles, Klein–Nishina physics also has important consequences for the scattered photons. When Comptonized deep in the Klein–Nishina limit, these may go on to pair produce with their parent population of ambient seed photons. Therefore, an astrophysically relevant treatment of reconnection with Klein–Nishina Compton cooling must also account for pair production.

So motivated, we provide, in this study, an analytical model for a relatively unexplored regime of radiative reconnection: the *pair-regulated Klein–Nishina regime*. Our model hinges on a self-regulation mechanism that we diagram in Fig. 1 and describe below. In that description as throughout this text, the term 'reconnection layer' (sometimes just 'layer') refers to the region permeated by reconnected magnetic flux; the term 'upstream' (sometimes 'inflow') refers to the region filled with unreconnected flux.

First (step 1 in Fig. 1), particles accelerated in the reconnection layer Comptonize ambient seed photons to gamma-ray energies.

Secondly (step 2 in Fig. 1), IC-produced gamma-rays penetrate into the upstream region about one pair-production mean free path from the layer. While propagating, these photons are immune to secondary IC scattering because the Thomson optical depth τ_{T} is very small (even though the pair-production optical depth $\tau_{\gamma\gamma}$ exceeds unity). In step 3, high-energy photons are absorbed by the background radiation, producing pairs in the upstream plasma. Newborn pairs are then advected towards the layer. While en route, they radiatively cool, and thus some of their initial energy never returns to the layer. Nevertheless, the created pairs remain hot enough that their energy density dominates that of the originally present colder upstream particles. Thus, the plasma feeding the layer in step 4 possesses a reduced magnetization – the ratio of magnetic energy density to total (original + hot pairs) matter enthalpy density. This inhibits particle acceleration and subsequent photon emission in the reconnection layer, closing the negative feedback loop.

Our model predicts two types of pair-regulated Klein–Nishina reconnection dynamics. If only a small fraction of the energy radiated away from the layer is recaptured as hot pairs, reconnection enters a steady state characterized by a *universal* (independent of the initial value) pair-regulated magnetization. However, if the nearly all of the radiated energy gets swept back into the layer as pairs, the steady state is never realized. Instead, the system overshoots its theoretical fixed point solution, getting caught between two extreme magnetization states. In these 'swing cycles', the high magnetization state yields efficient above-threshold photon emission from the layer and subsequent injection of hot pairs into the upstream region. This initiates a very low/pair-loaded magnetization. Here, pair production is quenched until the created pairs vacate the inflow plasma by entering the layer, restoring the high magnetization.

In both a steady state and a swing cycle, the created particles, when present, dominate the upstream pressure, but a prolific pair cascade is not expected. The (power-law) distribution of pairs injected into the inflow region, though potentially quite broad, is too steep for later pair generations – born from photons emitted by earlier upstream generations – to outnumber the first generation. Furthermore, for a wide range of parameters, the newborn pairs are also few in number relative to the original plasma particles.

This last aspect of our model qualitatively departs from earlier treatments of radiative reconnection with pair production (e.g. Lyubarskii 1996; Uzdensky 2011; Hakobyan et al. 2019). Rather than being dressed in a coat of pairs that dominates both the upstream matter energy and number densities, the reconnection layer in this regime is self-consistently fed by a few high-energy newborn particles that control only the energy, and hence the magnetization, of inflowing material.

Our model also predicts the power-law index of the particle energy distribution yielded by Klein–Nishina reconnection. However, this index is sensitive to our free parameters and simplifying assumptions. Therefore, although the present analysis holds promise for making contact with observed features (e.g. the photon power-law index) of FSRQs and black hole accretion disc coronae (ADCe), future numerical studies, for which this work lays the foundation, are necessary to refine the model and make testable predictions.

Our discussion is organized as follows. In Section 2, we review some of the salient features of radiative magnetic reconnection when low-energy Thomson IC scattering is the dominant radiative process. We focus here on organizing the relevant energy scales in the reconnection problem, classifying different radiative regimes from the hierarchy of these scales. This provides a jumping-off point from which to generalize our discussion to Klein–Nishina IC losses in Section 3. After introducing scales and classifying regimes of Klein–

Nishina reconnection in that section, we present our model, which pertains to a subset of the available scale hierarchies, in Section 4. We comment on the model's observable features and its implications for FSRQ jets and ADCe in the high-luminosity states of black hole X-ray binaries in Section 5. We conclude in Section 6.

2 REVIEW OF THOMSON RADIATIVE RECONNECTION

Although a variety of radiation mechanisms (e.g. synchrotron emission) may impact the dynamics of relativistic collisionless magnetic reconnection, in this work, we specialize to the case where external IC scattering is the dominant and most dynamically consequential radiative channel. In this section, we further restrict our discussion to the case where the IC process occurs in the low-energy Thomson limit (defined below). This allows us both to review some of the principal results applicable in this regime and to assemble a set of energy scales that will anchor our exploration of Klein–Nishina radiative reconnection in subsequent sections.

2.1 Single-particle Thomson IC cooling

To begin, we review Thomson IC radiative cooling at the level of individual particles. Afterwards, we extend the discussion to include collective effects governed by the interplay between radiation and reconnection.

We consider a static, homogeneous, isotropic bath of ambient radiation with spectral energy density $u_{ph}(\epsilon)$. An ultrarelativistic electron or positron with energy $E = \gamma m_e c^2 \gg m_e c^2$ traversing this radiation field preferentially IC-scatters photons that appear strongly blueshifted in its rest frame: the lab-frame photon energy ϵ transforms to the rest-frame energy $\epsilon' \sim \gamma\epsilon$ (except for a negligible few photons that travel within an angle $1/\gamma \ll 1$ of the particle's velocity and so transform to much smaller energies). If $\epsilon' \ll m_e c^2$, then the encounter reduces to Thomson scattering, which is approximately elastic, yielding final (scattered) rest-frame photon energy $\epsilon'_f \simeq \epsilon'$. Moreover, the Thomson differential cross-section lacks a strong angular dependence. Thus, very few photons are emitted to within angle $1/\gamma$ opposite the scattering particle's velocity, and the vast majority of scatterings yield typical final lab-frame photon energy $\epsilon_f \sim \gamma\epsilon'_f \sim \gamma^2\epsilon$.

The same assumptions that place us in the Thomson regime also imply that the particle loses a very small fraction of its energy in one collision: $\epsilon_f/\gamma m_e c^2 \sim \epsilon'/m_e c^2 \sim \gamma\epsilon/m_e c^2 \ll 1$. This allows IC radiative losses to be treated classically – as a continuous drag force, $\boldsymbol{f}_T = -(4/3)\sigma_T\gamma^2 U_{ph}\boldsymbol{\beta}$, where $\sigma_T = 8\pi e^4/3m_e^2 c^4$ is the Thomson cross-section, $U_{ph} \equiv \int d\epsilon\, u_{ph}(\epsilon)$ is the total background radiation energy density, and $c\boldsymbol{\beta}$ is the scattering particle's velocity vector (cf. Blumenthal & Gould 1970; Rybicki & Lightman 1979; Phinney 1982; Pozdnyakov, Sobol & Syunyaev 1983; Uzdensky 2016; Werner et al. 2019; Mehlhaff et al. 2020; Sironi & Beloborodov 2020). Importantly, $\boldsymbol{f}_T$ depends only on U_{ph} and not on the spectral distribution $u_{ph}(\epsilon)$ of IC seed photons. Thus, provided a plasma radiates purely in the Thomson IC regime and is also optically thin to Thomson scattering, the collective dynamics are insensitive to the incident spectrum (though the *Comptonized* spectrum is not). The Thomson IC radiated power per particle is

$$P_T(\gamma) = |c\boldsymbol{\beta} \cdot \boldsymbol{f}_T| = \frac{4}{3}\sigma_T c\beta^2\gamma^2 U_{ph}. \tag{1}$$

All of the features discussed here – continuous emission, $\epsilon_f \propto \gamma^2$, and incident-spectrum independence – change significantly when we later allow particles to experience general Klein–Nishina Compton losses. However, before we get there, let us move on from this single-particle picture of IC cooling and chart out how Thomson radiative cooling impacts the collective reconnection dynamics.

2.2 Thomson IC effects on collective plasma behaviour

We first introduce a set of parameters, cast as particle energy scales (Lorentz factors), that characterize Thomson IC scattering in the context of magnetic reconnection. We then examine the radiative regimes represented by the possible scale hierarchies.

2.2.1 Reconnection energy scales

The reconnecting magnetic field, B_0, can be recast in terms of a length-scale: the nominal relativistic gyroradius $\rho_0 \equiv m_e c^2/eB_0$. This is a useful form for comparing B_0 with other length-scales in the problem. For example, introducing the length of the reconnection layer L (see Fig. 1), one may define a 'Hillas criterion': the energy,

$$\gamma_{Hillas} \equiv \frac{eB_0 L}{m_e c^2} = \frac{L}{\rho_0}, \tag{2}$$

of a particle with Larmor radius $\gamma_{Hillas}\rho_0$ equal to the system size (Hillas 1984).[1] Equivalently, γ_{Hillas} is the energy imparted to a particle accelerated across the system by an electric field of strength B_0.

While γ_{Hillas} gives a firm upper bound on the achievable particle energies, a more practical scale for reconnection problems is

$$\gamma_{max} \equiv 0.1\gamma_{Hillas}, \tag{3}$$

corresponding to extreme acceleration (Aharonian et al. 2002; Uzdensky et al. 2011; Cerutti et al. 2012a). We use γ_{max} instead of γ_{Hillas} because the reconnection electric field, E_{rec}, is not quite as strong as B_0, but about equal to $\beta_{rec}B_0$ where β_{rec} is the dimensionless reconnection rate. For collisionless reconnection, β_{rec} can be expressed in terms of the Alfvén speed v_A as $\beta_{rec} \simeq 0.1v_A/c$. In the relativistic limit treated in this paper, $v_A \simeq c$ and hence $\beta_{rec} \simeq 0.1$.

It is sometimes helpful to equivalently define γ_{max} by equating time-scales. We therefore introduce the acceleration time for a particle being linearly accelerated by the electric field $E_{rec} \simeq 0.1B_0$:

$$t_X(\gamma) \equiv \frac{\gamma m_e c^2}{ec E_{rec}} = \frac{\gamma m_e c^2}{\beta_{rec}ec B_0} \simeq \frac{10\gamma\rho_0}{c} \simeq 1.6\frac{2\pi\gamma\rho_0}{c}, \tag{4}$$

equal to $10/2\pi \simeq 1.6$ cyclotron periods in the magnetic field B_0. Here, subscript 'X' denotes that the spatial regions where this type of linear acceleration is active usually surround magnetic X-points – locations where the magnetic field reconnects and near which particles become unmagnetized (e.g. Uzdensky et al. 2011; Cerutti et al. 2012a). Equivalently to (3), one may define γ_{max} such that a particle is X-point-accelerated over one system light crossing time:

$$t_X(\gamma_{max}) \equiv \frac{L}{c}. \tag{5}$$

Another important reconnecting system parameter is the (combined electron + positron) upstream plasma density n_0. With B_0 fixed, n_0 can be cast in terms of the (dimensionless) upstream cold magnetization

$$\sigma_{c,0} \equiv \frac{B_0^2}{4\pi n_0 m_e c^2}, \tag{6}$$

---

[1] For simplicity, we assume that L characterizes the size of the system in all spatial dimensions in addition to its horizontal length (Fig. 1).

Physically, $\sigma_{c,0}$ represents the initial magnetic energy per particle: $\sigma_{c,0} m_e c^2/2$. Since reconnection delivers an appreciable fraction of the magnetic energy to the plasma, $\sigma_{c,0}$ also characterizes the average Lorentz factor $\langle\gamma\rangle$ of reconnection-energized particles (absent radiative cooling). Assuming that half of the initial magnetic energy is dissipated, we have $\langle\gamma\rangle \sim \sigma_{c,0}/4$ (cf. Sironi, Petropoulou & Giannios 2015; Werner et al. 2016; Sironi & Beloborodov 2020). In this way, just like $\gamma_{\max}$ furnishes a characteristic particle energy scale to stand in for the system size L, the cold magnetization provides an energy scale that acts as a proxy for the upstream number density n_0.

We report here, for reference, one more important dimensionless parameter, the hot magnetization (cf. Melzani et al. 2014; Werner et al. 2018)

$$\sigma_{h,0} \equiv \frac{B_0^2}{4\pi w_0}. \tag{7}$$

Here, $w_0 = p_0 + u_0$ is the relativistic plasma enthalpy density with p_0 the upstream pressure and u_0 the internal energy density. For a relativistically hot upstream plasma with temperature $\theta_0 = T_0/m_e c^2 \gg 1$, the enthalpy density is $w_0 \simeq 4\theta_0 n_0 m_e c^2$ ($p_0 = \theta_0 n_0 m_e c^2 = u_0/3$); for a cold $\theta_0 \ll 1$ plasma, $w_0 \simeq u_0 \simeq n_0 m_e c^2$ is dominated by rest-mass energy. Thus, in the relativistically hot case, $\sigma_{h,0} \simeq 1/(2\beta_{\rm pl}) = \sigma_{c,0}/(4\theta_0) \ll \sigma_{c,0}$ where $\beta_{\rm pl} = 8\pi p_0/B_0^2$ is the plasma beta parameter, but in the opposite limit, $\sigma_{h,0} \simeq \sigma_{c,0}$. Physically, the hot magnetization determines whether the energy flux into the reconnection region is dominated by the magnetic field ($\sigma_{h,0} \gg 1$) or by the matter ($\sigma_{h,0} \ll 1$). In the former $\sigma_{h,0} \gg 1$ limit, the Alfvén speed, $v_A = c\sqrt{\sigma_{h,0}/(1+\sigma_{h,0})}$, approaches c and, as a result, magnetic reconnection may drive not only relativistic individual particle motion (which merely requires $\sigma_{c,0} \gg 1$) but also relativistic bulk flows. Thus, $\sigma_{h,0} \gg 1$ is really the defining feature of 'relativistic' reconnection. In the spirit of our present discussion, one may regard $\sigma_{h,0}$ as a proxy for the upstream temperature θ_0, especially when $\theta_0 \gg 1$. However, we do not make explicit use of $\sigma_{h,0}$ for some time (until Section 4). For now, we simply assume that we are in the relativistic limit of reconnection, $\sigma_{h,0} \gg 1$, and, within that context, scope out the possible regimes by ordering our other important energy scales (which so far include $\gamma_{\max}$ and $\sigma_{c,0}$).

Armed with $\gamma_{\max}$ and $\sigma_{c,0}$ (and also assuming $\sigma_{h,0} \gg 1$), we can describe the size of the relativistic reconnection system. A 'large' system satisfies $\gamma_{\max} \gg \sigma_{c,0}$. In terms of the global geometry, this implies a high aspect ratio: the system is much longer than the microscopic current sheet thickness, of order $\langle\gamma\rangle \rho_0$ – the typical Larmor radius of reconnection-energized particles (specifically, $\gamma_{\max} \gg \sigma_{c,0} \Rightarrow L \gg 10\sigma_{c,0}\rho_0$; cf. Werner et al. 2016). This renders the layer plasmoid-unstable and initiates plasmoid-dominated reconnection (e.g. Ji & Daughton 2011).

In terms of individual particles, $\gamma_{\max} \gg \sigma_{c,0}$ also alleviates system-size constraints on particle energization, at least up to the mean Lorentz factor $\langle\gamma\rangle \sim \sigma_{c,0}/4$. In addition, when $1 \ll \theta_0 \ll \langle\gamma\rangle \sim \sigma_{c,0}/4$ and, hence, $\sigma_{h,0} \gg 1$, previous non-radiative 2D particle-in-cell (PIC) simulations have found that direct/fast acceleration by the reconnection electric field saturates at $\gamma_X \sim$ (several)$\sigma_{c,0}$ (e.g. $4\sigma_{c,0}$; Werner et al. 2016; Kagan, Nakar & Piran 2018). Thus, when $\gamma_{\max} \gg \sigma_{c,0}$, both the physics governing average-energy particle acceleration and the direct acceleration mechanism are unencumbered by the system size.

However, we remark that there are, in addition to primary X-point acceleration, other energization channels in the large-system plasmoid-dominated regime: most of them take place on slower time-scales but are not limited to Lorentz factors of order several $\sigma_{c,0}$. We term these 'secondary' acceleration processes because they typically

operate on plasma that has already been processed into the layer (the region of reconnected magnetic flux). One example of such a process is adiabatic heating by slowly compressing reconnected magnetic fields inside plasmoids (Petropoulou & Sironi 2018; Hakobyan et al. 2021). A different, but related, example is a *Fermi*-type mechanism (Drake et al. 2006; Dahlin, Drake & Swisdak 2014; Guo et al. 2014, 2015, 2016, 2019, 2020) in which particles bounce from end to end across contracting plasmoids. (See Uzdensky 2020 for a review of secondary acceleration mechanisms in 2D reconnection.) The existence of such acceleration channels may, without radiative cooling, allow the highest particle energies to grow well beyond γ_X – even if the dynamics of average particles and X-point acceleration top out at much lower energies. This is one aspect where even weakly radiative reconnection differs from its non-radiative counterpart. Radiative losses can impose a system-size-independent high-energy cut-off even on acceleration processes with no intrinsic upper energy limit (e.g. Mehlhaff et al. 2020; Hakobyan et al. 2021), potentially allowing the spectrum of accelerated particles to become independent of L. With this in mind, we now turn to quantifying the impact of Thomson IC radiative cooling on reconnection.

As discussed above, particles emitting in the Thomson limit suffer a drag force $\boldsymbol{f}_{\rm IC}$ determined by the total background radiation energy density $U_{\rm ph}$. Thus, Thomson radiation introduces just one extra parameter, $U_{\rm ph}$, into the reconnection problem. To cast $U_{\rm ph}$ as an energy scale, we define the Lorentz factor, $\gamma_{\rm rad,T}$, at which radiative drag matches the acceleration force from the reconnection electric field, $eE_{\rm rec} \simeq 0.1eB_0$, or (equivalently) such that the X-point acceleration time, $t_X(\gamma)$, equals the Thomson cooling time, $t_{\rm cool,T}(\gamma) \equiv \gamma m_e c^2/P_T(\gamma)$, over which a particle radiates a significant fraction of its energy γ. Putting $f_{\rm IC}(\gamma_{\rm rad,T}) \equiv 0.1eB_0$ [or $t_X(\gamma_{\rm rad,T}) \equiv t_{\rm cool,T}(\gamma_{\rm rad,T})$] yields (cf. Nalewajko 2016; Uzdensky 2016; Werner et al. 2019; Mehlhaff et al. 2020; Sironi & Beloborodov 2020)

$$\gamma_{\rm rad,T} \equiv \sqrt{\frac{0.3eB_0}{4\sigma_T U_{\rm ph}}}. \tag{8}$$

There is also a third way to define $\gamma_{\rm rad,T}$: it is the Lorentz factor of a particle that cools in ~ 1 cyclotron period in the magnetic field B_0 (cf. Uzdensky 2016).

Despite the fact that it is defined only by balancing radiative cooling against X-point energization, the energy $\gamma_{\rm rad,T}$ firmly radiatively caps the achievable particle energies (Werner et al. 2019; Mehlhaff et al. 2020). This is because acceleration by the coherent reconnection electric field may be the fastest significant particle energization channel in magnetic reconnection.[2] Other channels (such as those discussed by Petropoulou & Sironi 2018; Guo et al. 2019, 2020; Hakobyan et al. 2021), while not saturating at γ_X like X-point acceleration, are much slower and hence radiatively stall at Lorentz factors less than $\gamma_{\rm rad,T}$ (e.g. Mehlhaff et al. 2020).

To sum up, we now have three energy scales, each characterizing different physical parameters in relativistic reconnection. Two of these, $\gamma_{\max}$ and $\sigma_{c,0}$ – representing, respectively, the system size L and the upstream particle density n_0 – are non-radiative, common to all reconnection problems. The third scale, $\gamma_{\rm rad,T}$, encodes the

---

[2]It is unclear whether motional electric fields $\boldsymbol{E}_{\rm motion} = -\boldsymbol{v} \times \boldsymbol{B}/c$ ($v \simeq v_A \simeq c$) – e.g. due to rapidly moving compact plasmoid cores where $B > B_0$ and, hence, $E_{\rm motion} \sim v_A B/c > B_0$ – yield an overall faster effective acceleration than X-point regions. A similar remark applies to momentary pulses of high electric fields associated with waves launched at plasmoid mergers (Philippov et al. 2019). In both cases, the issue is not just one of field strength but also of spatio-temporal coherence.

Table 1. Scale hierarchies and associated radiative regimes of relativistic reconnection subject to Thomson IC losses. A relativistic large-system ordering $\gamma_{\max} > \gamma_X > \langle\gamma\rangle > 1$ is assumed throughout. Inserting $\gamma_{\mathrm{rad,T}}$ and γ_{cool} at different locations in this base ordering corresponds to different regimes. To guide the eye, the 'independent' parameters (i.e. independent of $\gamma_{\max}$, γ_X, and $\langle\gamma\rangle$ but not of each other) $\gamma_{\mathrm{rad,T}}$ and γ_{cool} are typeset in red. The regimes are semantically distinguished based on which particles – either those in the bulk of the particle energy distribution near the average energy $\langle\gamma\rangle$ or those in the high-energy tail – cool on time-scales shorter than L/c (strong cooling) versus those that cool in one gyroperiod (saturated cooling). For Thomson IC radiation reaction (but not in the more general Klein–Nishina limit) saturated cooling is more efficient than, and therefore implies, strong cooling.

Scale hierarchy	Regime name	Bulk particles		High-energy particles	
		Strong cooling (Y/N)	Saturated cooling (Y/N)	Strong cooling (Y/N)	Saturated cooling (Y/N)
$\gamma_{\mathrm{cool}} > \gamma_{\mathrm{rad,T}} > \gamma_{\max} > \gamma_X > \langle\gamma\rangle$	Non-radiative	N	N	N	N
$\gamma_{\max} > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}} > \gamma_X > \langle\gamma\rangle$	Quasi non-radiative	N	N	Y^a	N
$\gamma_{\max} > \gamma_{\mathrm{rad,T}} > \gamma_X > \gamma_{\mathrm{cool}} > \langle\gamma\rangle$		N	N	Y	N
$\gamma_{\max} > \gamma_{\mathrm{rad,T}} > \gamma_X > \langle\gamma\rangle > \gamma_{\mathrm{cool}}$		Y	N	Y	N
$\gamma_{\max} > \gamma_X > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}} > \langle\gamma\rangle$		N	N	Y	Y
$\gamma_{\max} > \gamma_X > \gamma_{\mathrm{rad,T}} > \langle\gamma\rangle > \gamma_{\mathrm{cool}}$		Y	N	Y	Y
$\gamma_{\max} > \gamma_X > \langle\gamma\rangle > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}}$	Extremely radiative	Y	Y	Y	Y

^{a}Highest energy particles may or may not achieve strong cooling. If they do, it is not through impulsive X-point acceleration (see the text).

energy density of ambient radiation and is unique to reconnection with IC cooling. Furthermore, one of these energies, $\sigma_{c,0}$, splits into two: it represents both the average energy of reconnection-energized particles $\langle\gamma\rangle \sim \sigma_{c,0}/4$ and the intrinsic maximum energy deliverable by the X-point acceleration mechanism $\gamma_X \sim 4\sigma_{c,0}$ in the large-system plasmoid-mediated regime. We later argue (Section 5.1.2) that, under some circumstances, γ_X may exceed the nominal value $\gamma_X = 4\sigma_{c,0}$ found from 2D PIC simulations (Werner et al. 2016; Kagan et al. 2018), but in the case where γ_X truly is of order $4\sigma_{c,0}$, γ_X and $\langle\gamma\rangle$ are only offset by about a factor of 16.

2.2.2 Regimes of Thomson radiative reconnection

Next, we enumerate the various orderings of the scales discussed above and examine the physical regimes each ordering represents. To simplify this program, we concentrate solely on the large-system regime with a relativistic amount of magnetic energy per particle: $\gamma_{\max} \gg \sigma_{c,0} \gg 1$ (or, if we split the scale $\sigma_{c,0}$ into $\gamma_X \sim 4\sigma_{c,0}$ and $\langle\gamma\rangle \sim \sigma_{c,0}/4$, the regime $\gamma_{\max} \gg \gamma_X \gg \langle\gamma\rangle \gg 1$). Different scale hierarchies are then realized by inserting $\gamma_{\mathrm{rad,T}}$ into various positions of the base ordering $\gamma_{\max} \gg \gamma_X \gg \langle\gamma\rangle \gg 1$. The possible orderings are summarized in Table 1.

This procedure is made more conceptually transparent if we introduce the derived scale γ_{cool}, the Lorentz factor of a particle that cools in one dynamical time L/c of the system. Writing $t_{\mathrm{cool,T}}(\gamma_{\mathrm{cool}}) \equiv L/c$ yields

$$\gamma_{\mathrm{cool}} \equiv \gamma_{\mathrm{rad,T}}^2/\gamma_{\max}. \qquad (9)$$

Interestingly, the radiatively limited Lorentz factor $\gamma_{\mathrm{rad,T}}$ is always intermediate between γ_{cool} and $\gamma_{\max}$, equal to the geometric mean of those two scales. Note that one may have $\gamma_{\mathrm{cool}} < 1$, in which case γ_{cool} does not correspond to a physical Lorentz factor. In that case, all particles cool to non-relativistic energies in $t_{\mathrm{cool,T}}(1) < L/c$. Related to γ_{cool} is the compactness of the system

$$\ell \equiv \frac{U_{\mathrm{ph}}\sigma_T L}{m_e c^2} = \frac{3}{4\gamma_{\mathrm{cool}}}. \qquad (10)$$

The time for a particle to cool from any initial γ to $\gamma = 1$ is $\sim L/c\ell$.

We begin our exploration of the various radiative regimes by quantifying the *non-radiative* limit $\gamma_{\mathrm{cool}} > \gamma_{\mathrm{rad,T}} > \gamma_{\max} \gg \sigma_{c,0} \gg 1$. Hereafter, we do not list $\sigma_{c,0} \gg 1$ explicitly. We also only use '>' (not '$\gg$') symbols, with the understanding that all regimes become

more distinct when the corresponding scales are well separated. The regime $\gamma_{\mathrm{cool}} > \gamma_{\mathrm{rad,T}} > \gamma_{\max} > \sigma_{c,0}$ corresponds to the limit $U_{\mathrm{ph}} \to 0$ and, hence, $\gamma_{\mathrm{rad,T}} \to \infty$. Here, no particle radiates a significant fraction of its energy within one dynamical time L/c, effectively decoupling radiation from reconnection. This is the regime mentioned in the Introduction where magnetic reconnection can, in principle, be studied on its own and the radiative signatures calculated independently.

The first step up in radiative efficiency might be called the *quasi-non-radiative regime* $\gamma_{\max} > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}} > \gamma_X > \langle\gamma\rangle$. Here, primary X-point acceleration does not impart enough energy to particles so that they significantly radiate on one dynamical time L/c, let alone so that they achieve radiative saturation $\gamma_{\mathrm{rad,T}}$. Secondary acceleration channels, on the other hand, might be able to deliver particles to energies $\geq\gamma_{\mathrm{cool}}$ so that those particles radiate faster than the global time-scale. However, this depends on the detailed nature of each secondary acceleration process – whether any of them radiatively stall above γ_{cool} is not guaranteed.

As an example of a secondary acceleration mechanism relevant to the quasi-non-radiative regime, one may consider particles slowly energized inside of adiabatically compressing magnetic islands (also 'plasmoids'), as detailed by Petropoulou & Sironi (2018) and Hakobyan et al. (2021). The energy where radiative losses shut this process down is determined by matching the plasmoid compression time to the particle cooling time-scale. As reported by Hakobyan et al. (2021), this upper-limit energy is $\gamma_{\mathrm{sec}} \sim m_e c^2 \beta_{\mathrm{rec}}/w\sigma_T U_{\mathrm{ph}}$, where $w \simeq 0.1L$ is the size of the largest plasmoids formed by reconnection. [We ignore that smaller plasmoids compress faster and so may yield, for the smaller number of particles they contain, higher γ_{sec} (Hakobyan et al. 2021).] In effect, $\gamma_{\mathrm{sec}} \sim \ell^{-1} = (4/3)\gamma_{\mathrm{cool}}$. Thus, this secondary energization channel only barely accelerates some particles up to γ_{cool}, and most reach energies much less than this. Hence, considering only this secondary mechanism, the reconnection process can be regarded as marginally non-radiative, with only very few highest energy particles cooling in less than one dynamical time.

Increasing the cooling efficiency once more brings us to the first of several truly *radiative* regimes where radiation is dynamically important for at least some of the particles. Aptly naming these regimes is cumbersome because different particles can experience varying degrees of radiative efficiency. The high-energy particles, for example, may be rapidly cooled and the particles at the average energy $\langle\gamma\rangle$ cooled quite slowly. Therefore, we do not classify

these regimes globally, calling the entire system weakly or strongly radiative, but we refer to them based on which populations of particles cool on various time-scales.

We call particles *strongly cooled* if radiation reaction causes them to lose an appreciable fraction of energy in less than one dynamical time L/c – i.e. if their Lorentz factors exceed γ_{cool}. This agrees with typical notions of strong cooling in astrophysics, which indicate that radiative cooling occurs faster than some macroscopic system time-scale. Correspondingly, we call particles with $\gamma < \gamma_{\mathrm{cool}}$ *weakly cooled* (because they are not strongly cooled) and sometimes *non-radiative* (because they do not radiate appreciably in a dynamical time). The particles with much higher energies, close to $\gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}}$, we say exhibit *saturated cooling*: their Lorentz factors are radiatively saturated because intense emission prevents further energization ($t_{\mathrm{cool,T}} = t_{\mathrm{X}}$). Although, in the Thomson IC limit, particles undergoing saturated cooling radiate much more efficiently than just strongly cooled particles, this is not always true once Klein–Nishina effects come into play (see Section 3.2 and Fig. 5). Thus, we wish to avoid associating the saturated cooling regime with a term connoting excessively efficient or fast cooling (e.g. 'very strong cooling').

Using these terms, we see that the scale hierarchy $\gamma_{\mathrm{max}} > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{X}} > \gamma_{\mathrm{cool}} > \langle \gamma \rangle$ indicates that average (or 'bulk') particles are weakly radiative; most of them cool slower than L/c ($\langle \gamma \rangle < \gamma_{\mathrm{cool}}$). Meanwhile, at least some of the high-energy particles accelerated by the primary X-point channel are strongly cooled ($\gamma_{\mathrm{X}} > \gamma_{\mathrm{cool}}$). Even so, radiative losses are not so fast as to hinder direct X-point acceleration, with $t_{\mathrm{X}}(\gamma_{\mathrm{X}})$ faster than $t_{\mathrm{cool,T}}(\gamma_{\mathrm{X}})$ because $\gamma_{\mathrm{X}} < \gamma_{\mathrm{rad,T}}$. Thus, X-point acceleration (because it is intrinsically capped to below $\gamma_{\mathrm{rad,T}}$) – and, hence, all other (known) secondary energization channels (because they are slow) – cannot deliver particles up to the radiative saturation limit $\gamma_{\mathrm{rad,T}}$.

Permuting scales again by swapping the positions of γ_{cool} and $\langle \gamma \rangle$, we arrive in the regime $\gamma_{\mathrm{max}} > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{X}} > \langle \gamma \rangle > \gamma_{\mathrm{cool}}$. Here, most particles radiate strongly because $\gamma_{\mathrm{cool}} < \langle \gamma \rangle < \gamma_{\mathrm{X}}$. However, like in the previous regime, virtually no particles are expected to achieve radiative saturation ($\gamma_{\mathrm{X}} < \gamma_{\mathrm{rad,T}}$), and X-point acceleration, while unaffected by radiative cooling on the short time-scale on which it occurs [$t_{\mathrm{X}}(\gamma_{\mathrm{X}}) < t_{\mathrm{cool,T}}(\gamma_{\mathrm{X}})$], does produce particles of sufficiently high energies ($> \gamma_{\mathrm{cool}}$) to be strongly cooled.

Next we arrive at scale hierarchies where at least a few particles exhibit saturated cooling. One such domain is $\gamma_{\mathrm{max}} > \gamma_{\mathrm{X}} > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}} > \langle \gamma \rangle$. Here, some particles are promptly accelerated near X-points to the upper-limit energy $\gamma_{\mathrm{rad,T}} < \gamma_{\mathrm{X}}$, but the bulk particles, with $\langle \gamma \rangle < \gamma_{\mathrm{cool}}$, barely radiate even on global L/c time-scales. This is perhaps the most extreme example of how vastly different the cooling rates can be for different reconnection-energized particles. However, unless γ_{X} can substantially exceed its nominal $4\sigma_{\mathrm{c,0}}$ value, this regime may not be realized in astrophysical contexts. This is because, if $\gamma_{\mathrm{X}} = 4\sigma_{\mathrm{c,0}}$, then $\gamma_{\mathrm{X}}/\langle \gamma \rangle \simeq 16$, implying that $\gamma_{\mathrm{rad,T}}/\gamma_{\mathrm{cool}} < 16$, and, through equation (9), that $\gamma_{\mathrm{max}}/\gamma_{\mathrm{rad,T}} < 16$, whereas γ_{max}, $\gamma_{\mathrm{rad,T}}$, and γ_{cool} are each usually separated by several decades in astrophysical systems (see Section 5.2).

Moving to the last two possible orderings, we have $\gamma_{\mathrm{max}} > \gamma_{\mathrm{X}} > \gamma_{\mathrm{rad,T}} > \langle \gamma \rangle > \gamma_{\mathrm{cool}}$, in which the high-energy particles accelerated near X-points attain radiative saturation ($\gamma_{\mathrm{rad,T}} < \gamma_{\mathrm{X}}$) and the bulk particles are strongly cooled ($\gamma_{\mathrm{cool}} < \langle \gamma \rangle$). Finally, there is an *extremely radiative* regime, $\gamma_{\mathrm{max}} > \gamma_{\mathrm{X}} > \langle \gamma \rangle > \gamma_{\mathrm{rad,T}} > \gamma_{\mathrm{cool}}$. Here, IC losses firmly cap the acceleration of nearly all particles – not even the formal mean energy, $\langle \gamma \rangle$, available per particle can be attained – and should have dramatic effects on the large-scale reconnection dynamics (Uzdensky 2016). Table 1 summarizes the radiative reconnection regimes discussed in this section.

To complete our tour of the Thomson IC reconnection landscape, we review some of the previously identified physical effects that occur in these regimes. Several systematic PIC studies of Thomson radiative reconnection have been conducted in recent years, including those by Werner et al. (2019), Sironi & Beloborodov (2020), and Mehlhaff et al. (2020). There have also been radiative PIC studies of reconnection with strong synchrotron cooling (e.g. Cerutti et al. 2013, 2014a, b; Yuan et al. 2016) and QED effects (like pair production; Hakobyan et al. 2019; Schoeffler et al. 2019). In some cases, the qualitative features of reconnection regimes mediated by different radiative processes are similar, but, in this section, which is intended chiefly as a jumping-off point for our more general discussion of Klein–Nishina physics to come, we focus only on those effects studied within the context of Thomson IC losses.

The three radiative PIC studies conducted by Werner et al. (2019), Sironi & Beloborodov (2020), and Mehlhaff et al. (2020) explored a number of regimes outlined in this section. Werner et al. (2019) studied the effect of Compton losses on large-scale reconnection dynamics and on non-thermal particle acceleration (NTPA), exploring all the way from the non-radiative regime to that of fully saturated high-energy cooling and strong bulk cooling ($\gamma_{\mathrm{max}} > \gamma_{\mathrm{X}} > \gamma_{\mathrm{rad,T}} > \langle \gamma \rangle > \gamma_{\mathrm{cool}}$). They found that radiation steepens the high-energy part of the non-thermal tail of reconnection-accelerated particles but that the overall reconnection rate is virtually unaffected. Mehlhaff et al. (2020) focused on the angular distributions of high-energy particles, showing that particles approaching radiative saturation ($\gamma_{\mathrm{rad,T}} \lesssim \gamma_{\mathrm{X}}$) exhibit energy-dependent collimation in momentum space, forming narrow beams at the highest energies. Strong radiative losses thus appear to be an essential ingredient in mediating this 'kinetic beaming' effect, which was first discovered by Cerutti et al. (2012b). Numerically exploring the scenario first outlined by Beloborodov (2017), Sironi & Beloborodov (2020) focused primarily on the regime $\gamma_{\mathrm{max}} > \gamma_{\mathrm{X}} > \gamma_{\mathrm{rad,T}} > \langle \gamma \rangle > \gamma_{\mathrm{cool}}$ where the bulk particles are strongly cooled and the highest energy particles saturate at $\gamma = \gamma_{\mathrm{rad,T}}$. They found that, here, plasmoids are generally filled with cold plasma that has already released much of its energy through Compton losses. The plasma kinetic energy inside plasmoids is then dominated by bulk motion. This motion is also subject to radiative drag and, hence, is slower than in the non-radiative case. Sironi & Beloborodov (2020) also confirmed that the highest energy particles are accelerated near reconnection X-points and top out at Lorentz factors close to $\gamma_{\mathrm{rad,T}}$.

3 OVERVIEW OF KLEIN–NISHINA RADIATIVE RECONNECTION

We now generalize our discussion to Klein–Nishina IC losses, focusing first on single-particle cooling and then on collective effects.

3.1 Single-particle Klein–Nishina IC cooling

We begin with some guiding intuition. One can readily infer that the quadratic Thomson scaling of the scattered photon energy, $\epsilon_{\mathrm{f}} \sim \gamma^2 \epsilon$, must break down at some point: the particle cannot emit a photon of greater energy than its own $\gamma m_e c^2$ (ignoring the small initial energy ϵ). A new physical regime must take over when $\gamma^2 \epsilon$ becomes of order $\gamma m_e c^2$. At that point, the particle can no longer radiate continuously; it will lose an order-unity fraction of its energy in a single scattering event. Moreover, for even higher Lorentz factors, the photon energy can scale, at most, linearly with $\gamma m_e c^2$. The following analysis shows how these basic observations are borne out quantitatively.

At high energies, when the Thomson limit begins to break down, the seed photon energy ϵ becomes a dynamically important variable, influencing not just the spectrum of Comptonized photons, but also the power radiated by a particle $P_{IC}(\gamma)$. To simplify our treatment in the presence of this complication, we specialize to a monochromatic distribution of background radiation

$$u_{ph}(\epsilon) = U_{ph}\delta(\epsilon - \epsilon_{ph}). \tag{11}$$

To quantify the IC cooling domain, it is useful to define a critical Lorentz factor

$$\gamma_{KN} \equiv \frac{m_e c^2}{4\epsilon_{ph}} \tag{12}$$

and a Klein–Nishina parameter

$$q \equiv \frac{4\gamma\epsilon_{ph}}{m_e c^2} = \frac{\gamma}{\gamma_{KN}} \,. \tag{13}$$

Scattering particles suffer little recoil from individual photons when $q \ll 1$ ($\gamma \ll \gamma_{KN}$): IC radiation proceeds in the Thomson regime. The opposite, deep Klein–Nishina limit is when $q \gg 1$ ($\gamma \gg \gamma_{KN}$). The crossover point $q = 1$ ($\gamma = \gamma_{KN}$) corresponds to setting the maximum Thomson emission energy, $4\gamma^2\epsilon_{ph}$, equal to the Comptonizing particle energy, $\gamma m_e c^2$.

The Lorentz factor γ_{KN}, like $\gamma_{rad,T}$ for U_{ph} in the Thomson limit, is the fundamentally new energy scale introduced by Klein–Nishina physics. It serves as a proxy for the underlying physical parameter ϵ_{ph}. By ordering γ_{KN} with respect to our other fundamental energy scales discussed in the preceding section, $\gamma_{rad,T}$, $\sigma_{c,0}$, and γ_{max}, we can determine what new radiative regimes of reconnection are accessible once Klein–Nishina effects have been added to our physical framework.

Before embarking on that task, however, we focus purely on the radiative physics (ignoring collective plasma effects), to build our intuition for how individual particles experience IC losses. In the presence of the seed photon distribution (11), the IC power radiated by a single particle becomes

$$P_{IC}(\gamma) = P_T(\gamma) f_{KN}(\gamma/\gamma_{KN}), \tag{14}$$

This is the same as the Thomson expression (1) but modified by the dimensionless function of q (cf. Jones 1968; Nalewajko et al. 2018; Appendix A)

$$f_{KN}(q) = \frac{9}{q^3}\left[\left(\frac{q}{2} + 6 + \frac{6}{q}\right)\ln(1+q) \right.$$
$$\left. - \frac{1}{(1+q)^2}\left(\frac{11}{12}q^3 + 6q^2 + 9q + 4\right) - 2 + 2\,\mathrm{Li}_2(-q)\right], \tag{15}$$

where Li_2 is the dilogarithm. Fig. 2 displays $f_{KN}(q)$ together with its asymptotic large-argument limit

$$f_{KN}(q \gg 1) \simeq (9/2q^2)\left[\ln(q) - 11/6\right] \tag{16}$$

and its approximate form (e.g. Moderski et al. 2005)

$$f_{KN}(q) \simeq \frac{1}{(1+q)^{1.5}}, \tag{17}$$

which is roughly correct up to $q \sim 10^4$ (at $q \simeq 10^4$, the error reaches a factor of 3 and begins increasing rapidly). The function $f_{KN}(q)$ tends to unity as q becomes small, as required in the Thomson-limit $P_{IC}(\gamma) \to P_T(\gamma)$. For large arguments, $f_{KN}(q)$ falls off quadratically with a logarithmic correction: the scattering cross-section is suppressed in the deep Klein–Nishina limit.

Let us now examine the 'discreteness' of radiative losses as q increases through unity. To do so, we note that, similar to the radiated
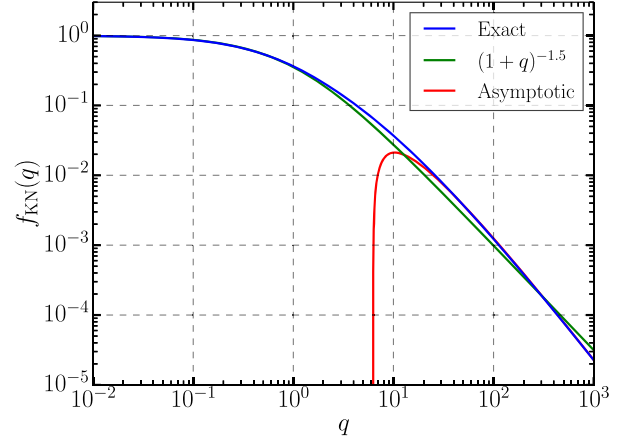

Figure 2. The function $f_{KN}(q)$ as defined in equation (15) in blue, its asymptotic form (16) in red, and its approximation (17) in green.

power $P_{IC}(\gamma)$, the rate $R_{IC}(\gamma)$ at which a single electron or positron scatters photons distributed according to (11) can be written as the corresponding Thomson rate, $R_T(\gamma) = R_T = c\sigma_T U_{ph}/\epsilon_{ph}$, times a dimensionless function:

$$R_{IC}(\gamma) = R_T\, g_{KN}(\gamma/\gamma_{KN}), \tag{18}$$

where $g_{KN}(q)$ reads (Appendix A)

$$g_{KN}(q) = \frac{3}{2q^2}\left[\left(q + 9 + \frac{8}{q}\right)\ln(1+q) \right.$$
$$\left. - \frac{1}{1+q}\left(\frac{q^2}{2} + 9q + 8\right) + 4\,\mathrm{Li}_2(-q)\right] \tag{19}$$

and has asymptotic form

$$g_{KN}(q \gg 1) \simeq \frac{3}{4q}\left[2\ln(q) - 1\right]. \tag{20}$$

Using (18), one can write down the average photon energy $\langle\epsilon\rangle(\gamma)$ emitted by a particle with Lorentz factor γ:

$$\langle\epsilon\rangle(\gamma) \equiv \frac{P_{IC}(\gamma)}{R_{IC}(\gamma)} = \left[\frac{\gamma}{3\gamma_{KN}}\frac{f_{KN}(\gamma/\gamma_{KN})}{g_{KN}(\gamma/\gamma_{KN})}\right]\gamma m_e c^2. \tag{21}$$

The quantity in square brackets here is the 'inelasticity' $\langle\epsilon\rangle(\gamma)/\gamma m_e c^2$, the typical fraction of a particle's energy lost in a single scattering encounter (Moderski et al. 2005). Using the small-argument limits $f_{KN}(q \ll 1) \to 1$ and $g_{KN}(q \ll 1) \to 1$, one may read off the well-known Thomson inelasticity

$$\lim_{\gamma \ll \gamma_{KN}}\frac{\langle\epsilon\rangle(\gamma)}{\gamma m_e c^2} = \frac{\gamma}{3\gamma_{KN}} \ll 1, \tag{22}$$

or, more commonly,

$$\lim_{\gamma \ll \gamma_{KN}}\langle\epsilon\rangle(\gamma) = \frac{\gamma^2 m_e c^2}{3\gamma_{KN}} = \frac{4}{3}\gamma^2\epsilon_{ph}. \tag{23}$$

Similarly, plugging in the asymptotic forms for f_{KN} and g_{KN} verifies that the inelasticity approaches unity as q is taken to infinity:

$$\lim_{\gamma \to \infty}\frac{\langle\epsilon\rangle(\gamma)}{\gamma m_e c^2} = \lim_{\gamma \to \infty}\frac{\ln(\gamma/\gamma_{KN}) - 11/6}{\ln(\gamma/\gamma_{KN}) - 1/2} = 1. \tag{24}$$

However, the limiting value is approached quite slowly, for the ratio in (24) is only $\simeq 1$ when $\ln(\gamma/\gamma_{KN}) \gg 11/6$. In fact, $\langle\epsilon\rangle(\gamma)$ does not surpass $3\gamma m_e c^2/4$ until $\gamma \gtrsim 300\gamma_{KN}$. Nevertheless, the inelasticity does obtain a value *of order* unity for much more modest q. For example, $\langle\epsilon\rangle(\gamma) \geq \gamma m_e c^2/4$ when $\gamma \gtrsim 2\gamma_{KN}$.
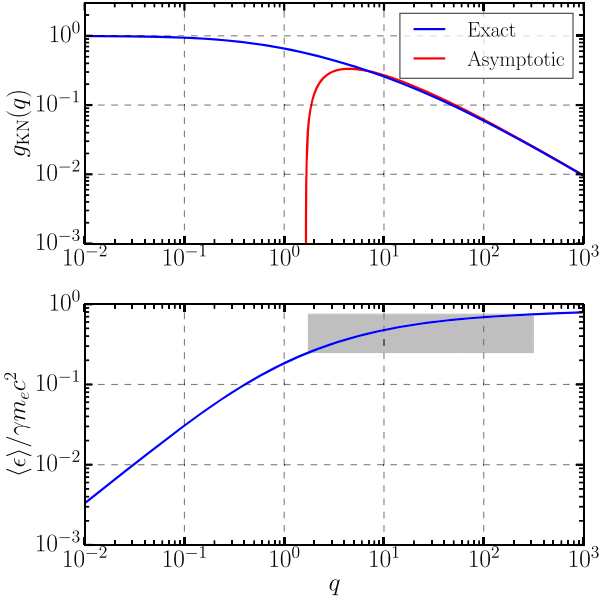
Figure 3. Top panel: The function $g_{KN}(q)$ (equation 19) in blue and its asymptotic form (20) in red. Bottom panel: The inelasticity $\langle\epsilon\rangle(\gamma)/\gamma m_e c^2 = q f_{KN}(q)/3 g_{KN}(q)$ (equation 21). The shaded region shows the appreciable range $q \in [1.75, 315]$ over which $\langle\epsilon\rangle$ is between $\gamma m_e c^2/4$ and $3\gamma m_e c^2/4$ (i.e. where it is close to $\gamma m_e c^2/2$). This indicates that, although $\langle\epsilon\rangle(\gamma)$ eventually approaches its asymptotic limit of $\gamma m_e c^2$, it does so rather slowly.

Both of these effects – the inelasticity slowly approaching, but rapidly rising to the vicinity of, unity for $q \gtrsim 1$ – as well as the function $g_{KN}(q)$ and its asymptotic form (20), are plotted in Fig. 3. In the figure, one sees that $\langle\epsilon\rangle(\gamma) \simeq \gamma m_e c^2/2$ for a wide range of q. Therefore, when we later need $\langle\epsilon\rangle$ for estimates, we adopt $\langle\epsilon\rangle(\gamma) \simeq \gamma m_e c^2/2$ rather than $\langle\epsilon\rangle \simeq \gamma m_e c^2$ whenever $q \gtrsim 1$. The latter becomes a more accurate approximation than the former for $q \gtrsim 300$, but it is unclear that the astrophysical systems we attempt to model contain particles at such high energies. (However, even if $q \gg 300$ in some systems, the distinction here between factors of order unity is well within the uncertainty of all of the estimates in this paper.) Furthermore, our model (Section 4) is mostly concerned with $q < 280$ (equation 60).

We have now verified our qualitative expectations for the deep Klein–Nishina regime: radiative losses become discrete when $q \gtrsim 1$, with particles losing an appreciable fraction of their energy to single photons. Moreover, the scattered photon energy scales approximately linearly with the pre-collision energy of the particle (rather than quadratically, as in the Thomson limit).

Next, because we are concerned primarily with IC radiation in this paper, we consider the circumstances in which synchrotron losses may be neglected. For an isotropic particle pitch-angle distribution, the average synchrotron power radiated per particle is

$$P_s = \frac{4}{3}\sigma_T c \beta^2 \gamma^2 U_B .\qquad(25)$$

Note that this is the same as the Thomson IC power (1) but with the ambient radiation energy density U_{ph} replaced by the magnetic field energy density $U_B \sim B_0^2/8\pi$ (we approximate the magnetic field strength throughout the reconnection system by its upstream value B_0). Equation (25) gives a total (IC + synchrotron) radiated

power per particle

$$P_{tot}(\gamma) = P_{IC}(\gamma) + P_s(\gamma)$$
$$= P_{IC}(\gamma)\left(1 + \frac{U_B}{U_{ph} f_{KN}(\gamma/\gamma_{KN})}\right).\qquad(26)$$

Clearly, IC losses dominate if

$$f_{KN}(q) > \frac{U_B}{U_{ph}} .\qquad(27)$$

Because $f_{KN}(q) \leq 1$, this criterion can only be met for systems whose ambient radiation energy density exceeds the magnetic field energy density. And, importantly, because Klein–Nishina effects begin to suppress IC cooling for $\gamma > \gamma_{KN}$, even when $U_{ph} \gg U_B$, there is always a high-energy Lorentz factor γ_s above which synchrotron losses dominate. Using the approximate form $f_{KN}(q) \simeq (1 + q)^{-1.5}$ (equation 17), one has (cf. Moderski et al. 2005)

$$\gamma_s \simeq \gamma_{KN}\left[\left(\frac{U_{ph}}{U_B}\right)^{2/3} - 1\right] .\qquad(28)$$

Thus, neglecting synchrotron losses is justified when γ_s exceeds the highest Lorentz factors reached in the system. We assume that this is indeed the case for the remainder of this work. We return to discuss the effects of finite γ_s as an effective limitation of our analysis in Section 5.

3.2 Klein–Nishina IC effects on collective plasma behaviour

We now examine how the collective reconnection dynamics are influenced by Klein–Nishina radiation reaction. We focus especially on differences from the case of purely Thomson radiative cooling.

Previously (in the Thomson regime), the highest Lorentz factor to which a particle could be accelerated was $\gamma_{rad,T}$. However, from equation (14), radiative losses are suppressed once q exceeds unity. This enables acceleration beyond $\gamma_{rad,T}$, and our definition of the radiative cut-off Lorentz factor can be generalized to include this effect. By equating the force from the reconnection electric field $0.1 e B_0$ to the (Klein–Nishina) Compton radiation reaction force $P_{IC}(\gamma)/c$, one may define a generalized cut-off $\gamma_{rad,IC}$ through

$$1 \equiv \frac{\gamma_{rad,IC}^2}{\gamma_{rad,T}^2} f_{KN}\left(\frac{\gamma_{rad,IC}}{\gamma_{KN}}\right) = \frac{\gamma_{rad,IC}^2}{\gamma_{rad,T}^2} f_{KN}\left(\frac{\gamma_{rad,IC}}{\gamma_{rad,T}}\frac{\gamma_{rad,T}}{\gamma_{KN}}\right) .\qquad(29)$$

The second equality explicitly shows that the ratio $\gamma_{rad,IC}/\gamma_{rad,T}$ is determined solely by the ratio $\gamma_{rad,T}/\gamma_{KN}$. Thus, $\gamma_{rad,IC}$ is a derived scale; it is fixed by the other radiative Lorentz factors γ_{KN} and $\gamma_{rad,T}$ (or, equivalently, through the physical parameters ϵ_{ph} and U_{ph}).

Equation (29) can always be satisfied for a finite $\gamma_{rad,IC}$, and a numerical solution to the equation is displayed in Fig. 4. When $\gamma_{rad,T} \ll \gamma_{KN}$, the equation is satisfied by $\gamma_{rad,IC} \simeq \gamma_{rad,T}$ because, in that case, $f_{KN}(\gamma_{rad,T}/\gamma_{KN} \ll 1) \simeq 1$. However, when $\gamma_{rad,T}$ becomes greater than γ_{KN}, the cut-off $\gamma_{rad,IC}$ becomes a rapidly increasing function of $\gamma_{rad,T}/\gamma_{KN}$. In fact, in the limit $\gamma_{rad,T} \gg \gamma_{KN}$, for which the large-argument approximation to f_{KN} (equation 16) applies, $\gamma_{rad,IC}$ grows superexponentially:

$$\gamma_{rad,IC} \simeq \gamma_{KN} \exp\left[\frac{2}{9}\left(\frac{\gamma_{rad,T}}{\gamma_{KN}}\right)^2 + \frac{11}{6}\right] .\qquad(30)$$

As shown in Fig. 4, this limiting form gives a good approximation to $\gamma_{rad,IC}$, even when $\gamma_{rad,T}$ only slightly exceeds γ_{KN}. We note that the rapid transition between the $\gamma_{KN} \gg \gamma_{rad,T}$ and $\gamma_{KN} \ll \gamma_{rad,T}$ limits is smoothed out in the presence of an extended (e.g. power-law) distribution of ambient radiation. Then, $\gamma_{rad,IC}$ does not grow
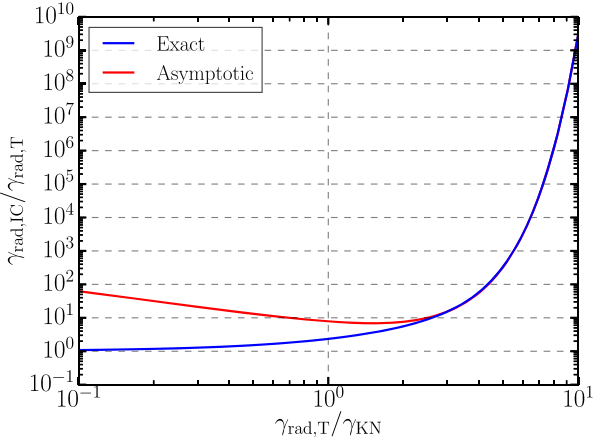
Figure 4. The Klein–Nishina cut-off Lorentz factor $\gamma_{\rm rad,IC}$ (29) and its asymptotic form (30), both normalized by the Thomson equivalent $\gamma_{\rm rad,T}$. The asymptotic form is already accurate when $\gamma_{\rm rad,T} \simeq 2.5\gamma_{\rm KN}$.

superexponentially until $m_e c^2/4\epsilon_{\rm low} \ll \gamma_{\rm rad,T}$, where $\epsilon_{\rm low}$ is the energy of the softest ambient photons.

Hence, when $\gamma_{\rm KN}$ becomes less than $\gamma_{\rm rad,T}$, Klein–Nishina physics inhibits radiative cooling from competing with rapid acceleration near reconnection X-points. It is then highly likely that the cut-off energy for X-point acceleration is set intrinsically rather than by radiative cooling (one expects $\gamma_{\rm X} \ll \gamma_{\rm rad,IC}$).

We now frame the superexponential divergence in $\gamma_{\rm rad,IC}$ from a different perspective: that of competing acceleration and radiative cooling time-scales. We then predict whether divergences occur in the cut-off energies of secondary acceleration channels by similarly comparing their time-scales against the cooling time. Including Klein–Nishina effects, the IC cooling time-scale is[3]

$$t_{\rm cool,IC}(\gamma) \equiv \frac{\gamma m_e c^2}{P_{\rm IC}(\gamma)} = \frac{\gamma_{\rm cool}}{\gamma f_{\rm KN}(\gamma/\gamma_{\rm KN})} \frac{L}{c}$$
$$= \frac{\gamma_{\rm cool}}{\gamma_{\rm KN}} [(\gamma/\gamma_{\rm KN}) f_{\rm KN}(\gamma/\gamma_{\rm KN})]^{-1} \frac{L}{c}, \qquad (31)$$

where $\gamma_{\rm cool}$ (equation 9) is the Lorentz factor of a particle with Thomson cooling time $t_{\rm cool,T}(\gamma) = \gamma m_e c^2/P_{\rm T}(\gamma)$ equal to L/c. A plot of $t_{\rm cool,IC}(\gamma)$ is presented in Fig. 5.

The relationship (31) encodes a wealth of information. In the Thomson regime $\gamma \ll \gamma_{\rm KN}$, $f_{\rm KN} \rightarrow 1$, and $t_{\rm cool,IC} \rightarrow t_{\rm cool,T} = \gamma_{\rm cool} L/\gamma c$, which decreases inversely with γ. However, for $\gamma \gg \gamma_{\rm KN}$, $f_{\rm KN}$ assumes its asymptotic form (16), inducing the scaling

$$t_{\rm cool,IC}(\gamma \gg \gamma_{\rm KN}) \simeq \frac{2L\gamma_{\rm cool}}{9c\gamma_{\rm KN}} \frac{\gamma/\gamma_{\rm KN}}{\ln(\gamma/\gamma_{\rm KN}) - 11/6}. \qquad (32)$$

Thus, in the deep Klein–Nishina limit, $t_{\rm cool,IC}(\gamma)$ *increases* linearly in γ with a logarithmic correction. This is the source of the superexponential divergence of $\gamma_{\rm rad,IC}$ with $\gamma_{\rm rad,T}$ in equation (30). Equation (4) shows that the time-scale, $t_{\rm X}$, on which X-point acceleration occurs is also linear in γ. Upon equating $t_{\rm cool,IC}$ and $t_{\rm X}$, which defines the radiative cut-off $\gamma_{\rm rad,IC}$, γ cancels. Thus, $t_{\rm X}$ can only surpass $t_{\rm cool,IC}$ if $\ln(\gamma/\gamma_{\rm KN})$ grows large enough, inducing the superexponential scaling in equation (30).

---

[3]As shown later (equation 36), the factor $\gamma_{\rm cool}/\gamma_{\rm KN}$ in equation (31) equals $3/5\tau_{\gamma\gamma}$ where $\tau_{\gamma\gamma}$ is the characteristic pair-production optical depth of the system.
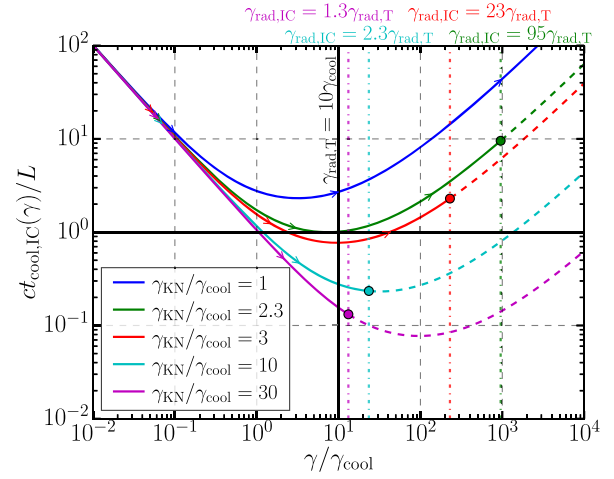
Figure 5. The IC cooling time $ct_{\rm cool,IC}(\gamma)/L$ (equation 31). In the plot, $\gamma_{\rm cool}$ is fixed, $\gamma_{\rm max} = 10\gamma_{\rm rad,T} = 100\gamma_{\rm cool}$, and $\gamma_{\rm KN}$ varies between curves. Arrows indicate the progression of particle cooling times during acceleration near reconnection X-points, which terminates at $\gamma = \gamma_{\rm rad,IC}$ (denoted by filled circles). For reference, each curve continues past $\gamma_{\rm rad,IC}$ as a dashed line. The cooling time is non-monotonic. Depending on $\gamma_{\rm KN}$, some (or even all) of the particles may attain high enough energies not to radiate [possessing cooling times $t_{\rm cool,IC}(\gamma) > L/c$]. If $\gamma_{\rm KN} < 2.3\gamma_{\rm cool}$, then all particles are non-radiative. If $\gamma_{\rm KN} > 2.3\gamma_{\rm cool}$, then only the low-energy ($\gamma < \gamma_{\rm cool,1} \sim \gamma_{\rm cool}$) and high-energy ($\gamma > \gamma_{\rm cool,2}$) particles are non-radiative. For particles to exceed $\gamma_{\rm cool,2}$, the condition $\gamma_{\rm rad,IC} \geq \gamma_{\rm cool,2}$ is necessary, but not sufficient (see the text).

Importantly, a similar situation does not arise for secondary acceleration channels, which are slower, possessing time-scales superlinear in γ (e.g. Mehlhaff et al. 2020). As an example, consider the secondary process described in Section 2, where particles inside contracting plasmoids are gradually energized. The Lorentz factors of such particles grow as $\gamma(t) \propto \sqrt{t}$ (Petropoulou & Sironi 2018; Hakobyan et al. 2021), yielding the acceleration time $t_{\rm sec}(t) = \gamma(t)/\dot\gamma(t) \propto \gamma^2(t)$. Thus, it is much easier for this secondary mechanism – and any other for which $t_{\rm sec} \propto \gamma^\zeta$ with $\zeta > 1$ – to radiatively saturate, as we now show.

For the sake of generality, suppose that $t_{\rm sec} = C\gamma^\zeta$ with C a constant independent of $\epsilon_{\rm ph}$ and $U_{\rm ph}$ (and, hence, of $\gamma_{\rm cool}, \gamma_{\rm rad,T}, \gamma_{\rm KN}$, and $\gamma_{\rm rad,IC}$; cf. the argument in Mehlhaff et al. 2020). Then, when IC cooling proceeds deep into the Klein–Nishina regime, the equality $t_{\rm cool,IC} = t_{\rm sec}$ reduces to $\gamma_{\rm sec}^{\zeta-1} \propto \gamma_{\rm cool}/\gamma_{\rm KN}^2$. The key difference from the direct X-point acceleration channel is that here we can ignore the $\ln(\gamma)$ correction – its dependence on γ is much weaker than $\gamma^{\zeta-1}$. As a result, the cut-off $\gamma_{\rm sec}$ scales merely polynomially in $\gamma_{\rm cool}$ (i.e. in $\gamma_{\rm rad,T}$) and $\gamma_{\rm KN}$: $\gamma_{\rm sec} \propto (\gamma_{\rm cool}/\gamma_{\rm KN}^2)^{1/(\zeta-1)}$. Plugging in $\zeta = 2$ for the adiabatic plasmoid compression process gives $\gamma_{\rm sec} \propto \gamma_{\rm cool}/\gamma_{\rm KN}^2$ (different from $\gamma_{\rm sec}$ reported in Section 2 because we are now considering deep Klein–Nishina cooling). Thus, Klein–Nishina physics may effectively remove the high-energy radiation reaction cap on impulsive X-point acceleration, but *not* on other processes, potentially increasing the relative importance of the primary direct energization channel.

As discussed above, $t_{\rm cool,IC}(\gamma)$ is non-monotonic, decreasing with γ when $\gamma \ll \gamma_{\rm KN}$ and increasing when $\gamma \gg \gamma_{\rm KN}$. It reaches the minimum

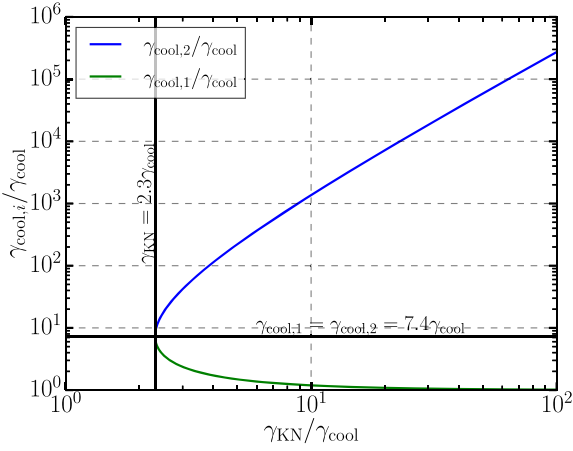$$\min_\gamma \left(\frac{ct_{\rm cool,IC}(\gamma)}{L} \right) \simeq 2.32 \frac{\gamma_{\rm cool}}{\gamma_{\rm KN}} \qquad (33)$$

Figure 6. The Lorentz factors $\gamma_{cool,1}$ and $\gamma_{cool,2}$ where $t_{cool,IC}(\gamma)$ dips below and above one dynamical time L/c, respectively. Since $\gamma_{cool,1}$ occurs almost entirely in the Thomson regime, it is almost always quite close to γ_{cool}, differing from γ_{cool} by at most the factor $7.4 = 3.20 \times 2.32$ (equations 33 and 34) when γ_{KN} approaches the non-radiative value $2.32\gamma_{cool}$. In contrast, $\gamma_{cool,2}$ grows approximately quadratically in $\gamma_{KN}/\gamma_{cool}$ (with a logarithmic correction).

at a critical *fastest-cooling Lorentz factor*

$$\arg\min_\gamma \left(\frac{ct_{cool,IC}(\gamma)}{L} \right) \simeq 3.20\gamma_{KN}, \tag{34}$$

[i.e. $\min_\gamma(t_{cool,IC}(\gamma)) = t_{cool,IC}(3.20\gamma_{KN})$]. The minimum cooling time (33) implies that, when $\gamma_{KN} < 2.32\gamma_{cool}$, *all* of the particles in the system radiate weakly (they have cooling times exceeding L/c). Even if γ_{KN} falls above this threshold and, hence, $\min(t_{cool,IC}) < L/c$, some high-energy particles may radiate weakly. Namely, if a particle surpasses the fastest-cooling Lorentz factor $3.20\gamma_{KN}$ by a sufficient amount, it reaches a high-energy domain with $t_{cool,IC}(\gamma) > L/c$. This effect does not occur in the Thomson regime.

These remarks are illustrated in Fig. 5. The figure displays $ct_{cool,IC}(\gamma)/L$ for fixed $\gamma_{cool} = \gamma_{rad,T}/10 = \gamma_{max}/100$ and several γ_{KN}. On each curve for which $\gamma_{KN} > 2.32\gamma_{cool}$, the line $t_{cool,IC} = L/c$ is crossed twice, once at a low Lorentz factor $\gamma_{cool,1}$ and once at a high Lorentz factor $\gamma_{cool,2}$. We analyse $\gamma_{cool,1}$ and $\gamma_{cool,2}$ shortly, but we point out some basic features of Fig. 5 beforehand. First, when $\gamma_{rad,IC} \simeq \gamma_{rad,T} \leq \gamma_{KN}$ (implying $\gamma_{KN} > 2.32\gamma_{cool}$ because $\gamma_{KN} > \gamma_{rad,T} = 10\gamma_{cool} > 2.32\gamma_{cool}$), particles may access only the Thomson portion of a cooling curve where $t_{cool,IC} \propto \gamma^{-1}$. The $\gamma_{KN} = 30\gamma_{cool} = 3\gamma_{rad,T}$ case illustrates this. Next, in the opposite limit, when γ_{KN} becomes smaller than $\gamma_{rad,T}$, the radiative cut-off $\gamma_{rad,IC}$ begins to grow rapidly, opening up the portion of a curve that bends upward. Eventually, at the critical Lorentz factor $\gamma_{cool,2}$, the cooling time $t_{cool,IC}$ once again equals L/c. Thus, if $\gamma_{rad,IC} > \gamma_{cool,2}$, particles accelerated near X-points could break into the high-energy weakly radiative regime. However, in reality, whether particles will actually cross this boundary does not depend solely on whether $\gamma_{rad,IC}$ surpasses $\gamma_{cool,2}$. That is just a necessary condition. In addition, the intrinsic X-point acceleration Lorentz factor γ_X must exceed $\gamma_{cool,2}$, or – if it does not – secondary acceleration channels must be able to energize particles against radiative cooling past $\gamma_{cool,2}$.

Both $\gamma_{cool,1}$ and $\gamma_{cool,2}$ are illustrated in Fig. 6 as functions of $\gamma_{KN}/\gamma_{cool}$ (they only depend on $\gamma_{rad,T}$ and γ_{max} through γ_{cool}). In general, $\gamma_{cool,1}$ is close to γ_{cool}, because, as illustrated in Fig. 5, particles are almost completely in the Thomson limit when $t_{cool,IC}$ crosses L/c from above. In contrast, $\gamma_{cool,2}$, where $t_{cool,IC}$ crosses L/c in the opposite direction, depends rather strongly on γ_{KN}. Be-

cause $\gamma_{cool,2}$ occurs fairly deep into the Klein–Nishina regime, one may employ expression (32) to see that $\gamma_{cool,2}$ satisfies

$$\frac{9\gamma_{KN}}{2\gamma_{cool}} \simeq \frac{\gamma_{cool,2}/\gamma_{KN}}{\ln(\gamma_{cool,2}/\gamma_{KN}) - 11/6}, \tag{35}$$

implying that $\gamma_{cool,2}/\gamma_{cool} \propto (\gamma_{KN}/\gamma_{cool})^2 \times \mathcal{O}(\ln(\gamma_{cool,2}/\gamma_{KN}))$.

Thus, $\gamma_{cool,2}$ is approximately quadratic in γ_{KN}. This differs from $\gamma_{rad,IC}$, which grows as γ_{KN} is *reduced*. Let us imagine that one starts with γ_{KN} high enough that $\gamma_{cool,2} \gg \gamma_{max} \gg \gamma_{rad,IC}$ (e.g. as on the $\gamma_{KN} = 30\gamma_{cool}$ curve in Fig. 5). Then, dialing down γ_{KN}, eventually $\gamma_{cool,2}$ will cross $\gamma_{rad,IC}$ and γ_{max} from above. It turns out that, by definition, these crossings occur simultaneously. For, if $\gamma_{cool,2} = \gamma_{max}$, then $t_X(\gamma_{max}) \equiv L/c$ (equation 5) and $t_{cool,IC}(\gamma_{cool,2}) \equiv L/c$ (equation 31). Consequently, $t_{cool,IC}(\gamma_{cool,2}) = t_X(\gamma_{cool,2})$, which implies $\gamma_{max} = \gamma_{cool,2} = \gamma_{rad,IC}$. By similar reasoning, one can show that, if $\gamma_{cool,2} = \gamma_{rad,IC}$, then $\gamma_{cool,2} = \gamma_{max}$. Continuing to reduce γ_{KN} beyond this 'triple point' yields the scale ordering $\gamma_{rad,IC} \gg \gamma_{max} \gg \gamma_{cool,2}$ (typified in the $\gamma_{KN} = 3\gamma_{cool}$ curve of Fig. 5, although the separation of scales is rather small in that example). The first of these inequalities, $\gamma_{rad,IC} \gg \gamma_{max}$, means that no particles achieve radiative saturation; all have Lorentz factors $\gamma < \gamma_{rad,IC}$ (see Section 2 discussion). The second inequality $\gamma_{cool,2} \ll \gamma_{max}$, potentially allows various acceleration channels (e.g. X-point acceleration if $\gamma_X > \gamma_{cool,2}$) to break a population of high-energy particles through the weakly radiative $\gamma = \gamma_{cool,2}$ barrier.

To summarize up to this point, we have encountered several stark departures from the Thomson picture of radiative cooling induced by finite γ_{KN}. Not only do particles with Lorentz factors $\gamma > \gamma_{KN}$ radiate their energy in discrete chunks, but their radiative cooling times can actually be quite long. When $\gamma_{KN} \lesssim \gamma_{rad,T}$, the effective cut-off Lorentz factor $\gamma_{rad,IC}$ begins to grow rapidly, and comes, with just a small change in γ_{KN}, to exceed γ_{max}, the analogue of the Hillas Lorentz factor for relativistic magnetic reconnection. When γ_{KN} is decreased even more, eventually it falls below γ_{cool}, and no particles in the system radiate efficiently. Thus, even in a nominally strongly radiative Thomson scale ordering, by making γ_{KN} small enough, a non-radiative regime can be reached. At intermediate γ_{KN}, a variety of intriguing and exotic physical effects can occur, which we elucidate later in this study. But first we cover one additional piece of physics that is entirely new to the Klein–Nishina realm: pair production.

3.2.1 Pair production in Klein–Nishina reconnection

A Comptonized photon of energy ϵ may collide with a background photon (energy ϵ_{ph}) to produce an electron–positron pair if the threshold criterion, $\epsilon\epsilon_{ph} \geq (m_ec^2)^2$, is met. In this work, we assume that $\gamma_{KN} = m_ec^2/4\epsilon_{ph} \gg 1$ ($\epsilon_{ph} \ll 100$ keV), implying that $\epsilon \gg m_ec^2 \gg \epsilon_{ph}$ is required to reach pair-production threshold. Such a high-energy photon can only be emitted in the Klein–Nishina IC regime. (The first author would here like to acknowledge Benoît Cerutti, who originally pointed this out to him.) If one assumes that the IC scattering occurs in the Thomson limit, then a contradiction arises because $1 \leq \epsilon\epsilon_{ph}/(m_ec^2)^2 \sim (\gamma\epsilon_{ph}/m_ec^2)^2 \sim q^2$. In contrast, assuming a Klein–Nishina scaling $\epsilon \sim \gamma m_ec^2/2$ yields the self-consistent result, $q \equiv 4\gamma\epsilon_{ph}/m_ec^2 \sim 8\epsilon\epsilon_{ph}/(m_ec^2)^2 \geq 8$. We adopt $\gamma_{pp} \equiv 8\gamma_{KN}$ (used mainly in Section 4) as the characteristic minimum particle Lorentz factor to emit above-threshold photons.

Now, the pair-production cross-section $\sigma_{\gamma\gamma}$ is zero precisely at threshold, $\epsilon\epsilon_{ph} = (m_ec^2)^2$, but, for the (isotropic, monochromatic) background distribution (11), it soon peaks at $\sigma_{\gamma\gamma} \simeq \sigma_T/5$ when $\epsilon\epsilon_{ph} \simeq 3.6(m_ec^2)^2$ (Gould & Schréder 1967). For such photons $q \sim 3.6 \times 8$
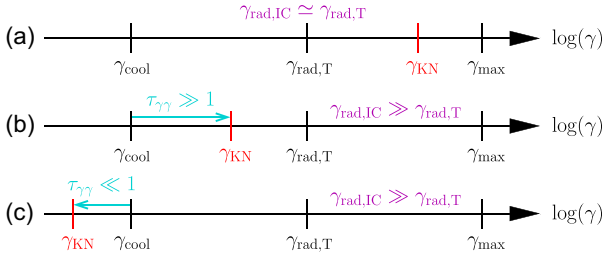
Figure 7. Basic orderings of γ_{KN} with respect to γ_{cool}, $\gamma_{rad,T}$, and γ_{max}. Note that $\gamma_{rad,T} \equiv (\gamma_{cool}\gamma_{max})^{1/2}$ (equation 9). The case $\gamma_{max} < \gamma_{rad,T} < \gamma_{cool}$, which is non-radiative irrespective of γ_{KN}, is not considered. The pair-production optical depth $\tau_{\gamma\gamma} \sim \gamma_{KN}/\gamma_{cool}$ is roughly the separation between γ_{KN} and γ_{cool}. The cut-off $\gamma_{rad,IC}$ is virtually identical to $\gamma_{rad,T}$ when $\gamma_{rad,T} \ll \gamma_{KN}$; otherwise, it is much larger than $\gamma_{rad,T}$. Case A: A Thomson radiative reconnection ordering (Section 2); $\gamma_{rad,T} \ll \gamma_{KN}$ prevents particles from accessing the Klein–Nishina regime and from emitting above-threshold photons. Case B: A scale hierarchy where Klein–Nishina IC losses (because $\gamma_{KN} \ll \gamma_{rad,T}$) and pair production (because $\gamma_{cool} \ll \gamma_{KN}$ implies $\tau_{\gamma\gamma} \gg 1$) are both likely. Case C: All orderings with $\gamma_{KN} \ll \gamma_{cool}$ are non/weakly radiative because $t_{cool,IC} > L/c$. Even the few photons radiated above pair threshold are not absorbed because $\tau_{\gamma\gamma} \ll 1$.

$\simeq 30$. Hence, although the energy scales at which Klein–Nishina IC cooling and pair production occur are both set, fundamentally, by γ_{KN}, they are offset from one another by a factor of about $10-30$. The former kicks in when $q \sim 1$ and the latter when $q \sim 8-30$. In this sense, the energy scale γ_{KN} 'splits', similarly to $\sigma_{c,0}$ (Section 2), into two that are offset by a fixed ratio.

However, it is not clear that astrophysical reconnection accelerates particles to energies that are high enough to stray from the Thomson limit but not to emit pair-producing photons. If, contrary to our simplified monochromatic assumption, the seed photons have any spread in energy, then photons Comptonized from the high-energy end of the background will more easily pair produce with the lower energy component. Even for a thermal radiation bath, the two frequencies where the Planck spectrum attains half its maximum value are offset from each other by about a factor of 5, reducing the effective splitting of γ_{KN} from a factor of $10-30$ to $2-6$. The cut-off in the reconnection-energized particle distribution would then have to fall precisely in a narrow range for Klein–Nishina effects to kick in but for pair production to remain impossible. And, even in this case, only the very highest energy sliver of particles would experience Klein–Nishina IC losses; most of the particles would still be cooled in the Thomson regime. Thus, from here on, we assume that Klein–Nishina IC scattering coincides with the emission of above-threshold photons in reconnection.

However, just because a high-energy photon is above threshold does not mean that it gets absorbed inside the reconnection system. One must also consider the optical depth, $\tau_{\gamma\gamma} = U_{ph}\sigma_{\gamma\gamma}L/\epsilon_{ph}$, to pair production. For simplicity, we evaluate $\tau_{\gamma\gamma}$ at the peak cross-section $\sigma_{\gamma\gamma} \simeq \sigma_T/5$, which is attained when $\epsilon\epsilon_{ph} \simeq 3.6 m_e c^2$. Thus,

$$\tau_{\gamma\gamma} \equiv \frac{U_{ph}\sigma_T L}{5\epsilon_{ph}} = \frac{3}{5}\frac{\gamma_{max}\gamma_{KN}}{\gamma_{rad,T}^2} = \frac{3}{5}\frac{\gamma_{KN}}{\gamma_{cool}}. \quad (36)$$

We have already encountered $\tau_{\gamma\gamma}$. It is the (inverse of the) prefactor in the expression for the cooling time $t_{cool,IC}(\gamma)$ in equation (31). Thus, the condition $\gamma_{KN} > 2.32\gamma_{cool}$ (equation 33), which ensures that at least some particles cool in times shorter than L/c, is the same as the optically thick condition $\tau_{\gamma\gamma} > (3/5) \times 2.32 \simeq 1$.

This means that there is an appreciable range of parameters where one expects both dynamically important Klein–Nishina radiative

cooling and pair production. Both mechanisms may actively feedback on the reconnection process when $\gamma_{cool} \ll \gamma_{KN} \ll \gamma_{rad,T}$. The first relationship, $\gamma_{cool} \ll \gamma_{KN}$, is necessary both for $\tau_{\gamma\gamma} \gg 1$ and for $\min[t_{cool,IC}(\gamma)] \ll L/c$. The second criterion $\gamma_{KN} \ll \gamma_{rad,T}$ is required for at least some particles to enter the regime where Klein–Nishina effects begin to impact their radiative cooling, also enabling them to emit photons above pair threshold.

We are thus equipped with a simple rule for deciding when Klein–Nishina and pair-production physics become important in reconnection. We just assemble all of our energy scales: γ_{max}, $\sigma_{c,0}$ (i.e. γ_X and $\langle\gamma\rangle$), $\gamma_{rad,T}$, and γ_{cool}, arrange these into a familiar Thomson hierarchy (as in Section 2), and insert γ_{KN} into a relevant location. If γ_{KN} is larger than $\gamma_{rad,T}$, Klein–Nishina effects are absent because $\gamma_{rad,T}$ imposes a hard upper bound on particle acceleration, and, consequently, no particles ever reach γ_{KN}. If, on the other hand, $\gamma_{KN} < \gamma_{cool}$, then Klein–Nishina effects suppress cooling so much that the whole system becomes non-radiative. Only if $\gamma_{cool} \ll \gamma_{KN} \ll \gamma_{rad,T}$ are Klein–Nishina IC cooling and pair production both important. And, in that case, it is also necessary to consider how $\gamma_{rad,T}$, γ_{KN}, and γ_{cool} are ordered with respect to the other scales in the problem. These remarks are illustrated in Fig. 7 and elaborated in the next subsection.

3.2.2 Regimes of Klein–Nishina radiative reconnection

We now systematically explore, as we did for Thomson IC cooling in Section 2.2.2, how to classify regimes of Klein–Nishina radiative reconnection. As an example, consider the Thomson ordering $\gamma_{max} > \gamma_{rad,T} > \gamma_X > \langle\gamma\rangle > \gamma_{cool}$ (4th row in Table 1). If we insert γ_{KN} between $\gamma_{rad,T}$ and γ_X, then Klein–Nishina effects do not affect primary X-point acceleration. They only come into play if secondary energization channels can push particles up to $\gamma \sim \gamma_{KN}$. If γ_{KN} is instead placed between γ_X and $\langle\gamma\rangle$, Klein–Nishina radiative cooling definitely impacts high-energy particles accelerated near X-points, and these particles are also likely to emit pair-producing photons. Klein–Nishina and pair-production physics become even more important if γ_{KN} is made smaller than $\langle\gamma\rangle \sim \sigma_{c,0}/4$. Then, the bulk of the accelerated particles – not just the high-energy tail – emit in the Klein–Nishina regime and, likely, many pairs are produced.

Because γ_{KN} is not the only new scale, but also introduces a few derived scales (e.g. $\gamma_{cool,2}$ and $\gamma_{rad,IC}$), exhaustively discussing all possible regimes like in Section 2.2.2 is prohibitively tedious. Even in the preceding paragraph, we did not consider subtleties such as whether $\gamma_X > \gamma_{cool,2}$, in which case some high-energy particles radiate inefficiently. In lieu of an exhaustive discussion, we supply Fig. 8, a 'phase diagram', in the $\gamma_{rad,T}/\sigma_{c,0}-\gamma_{KN}/\sigma_{c,0}$ plane, of the complex radiative parameter space for Klein–Nishina reconnection. The parameter space is, in reality, 3-dimensional, depending also on $\gamma_{max}/\sigma_{c,0}$. To display it in 2D, we set $\gamma_{max} = 10^3\sigma_{c,0}$ in Fig. 8.

In the figure, contours highlight important values of energy scales and often distinguish different-coloured regimes of interest. We caution that the colour scheme is somewhat arbitrary. Almost every sliver of parameter space enclosed within a set of contours is its own physical regime, and only a subset of the relevant contours are shown. Without analysing every possible contour-enclosed region, the best we can do is group regions based on similar expected qualitative behaviour, a heuristic that guides the colour-coding in Fig. 8. However, this exercise is ultimately subjective. A given grouping is useful for conceptualizing some physical similarities, but may need to be reevaluated if the physics of main interest changes.
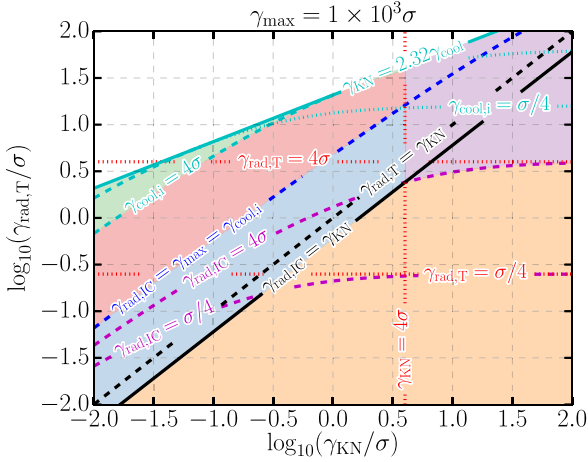
Figure 8. Radiative phase diagram for Klein–Nishina relativistic magnetic reconnection. The diagram is inherently three-dimensional – it depends on $\gamma_{\rm rad,T}/\sigma_{\rm c,0}$, $\gamma_{\rm KN}/\sigma_{\rm c,0}$, and $\gamma_{\rm max}/\sigma_{\rm c,0}$ – but we display a 2D slice where $\gamma_{\rm max} = 10^3\sigma_{\rm c,0}$. In the plot, $\sigma_{\rm c,0}$ is abbreviated as σ. Contours show important equalities between energy scales. The $\gamma_{\rm cool,i} = {\rm const.}$ contours transition, when they strike the curve $\gamma_{\rm KN} = 2.32\gamma_{\rm cool}$, from a dashed portion along which $\gamma_{\rm cool,2}$ is fixed, to a dotted portion where $\gamma_{\rm cool,1}$ is fixed to the same value. Regions with a common colour exhibit similar qualitative behaviour. The colours are inspired by assuming that $\gamma_{\rm X} = 4\sigma_{\rm c,0}$ and, fictitiously, that X-point acceleration is the only operative energization mechanism. Thus, the highest Lorentz factor attained is $\min(\gamma_{\rm rad,IC}, 4\sigma_{\rm c,0})$. In the purple region, $4\sigma_{\rm c,0} < \gamma_{\rm KN}, \gamma_{\rm rad,IC}$. Hence, radiative cooling does not furnish an upper bound on particle acceleration, and no particles suffer Klein–Nishina losses. In the orange region, $\gamma_{\rm rad,IC} < \gamma_{\rm KN}, 4\sigma_{\rm c,0}$. Here, Thomson IC losses firmly cap the maximum particle energy to $\gamma_{\rm rad,IC} \simeq \gamma_{\rm rad,T} < \gamma_{\rm KN}$, and, consequently, the highest energy particles are radiatively saturated. In the blue region, $\gamma_{\rm KN} < \gamma_{\rm X}, \gamma_{\rm rad,IC}$ and $\gamma_{\rm rad,IC} < \gamma_{\rm max} < \gamma_{\rm cool,2}$. Because $\gamma_{\rm KN} < \gamma_{\rm X}$ some particles reach $q \geq 1$. In addition, in the subset of the region where $\gamma_{\rm rad,IC} < \gamma_{\rm X} = 4\sigma_{\rm c,0}$, the highest energy particles are radiatively saturated. The red region is defined by $\gamma_{\rm rad,IC} > \gamma_{\rm max} > \gamma_{\rm cool,2} > 4\sigma_{\rm c,0}$. Thus, radiative losses do not bound particle acceleration, and, as a result, there is likely a broad spectrum of particles above the threshold energy $\gamma \simeq 8\gamma_{\rm KN}$ to emit pair-producing radiation. In the green $\gamma_{\rm cool,2} < 4\sigma_{\rm c,0}$ region, Klein–Nishina effects suppress cooling so that some high-energy particles are weakly radiative, having $\gamma > \gamma_{\rm cool,2}$ and, hence, $t_{\rm cool,IC} > L/c$. In the white $\gamma_{\rm KN} < 2.32\gamma_{\rm cool}$ region, all particles have $t_{\rm cool,IC} > L/c$.

We now describe the (colour-coded) grouping of regions adopted in Fig. 8. We begin with the fundamentally new domains corresponding to case B in Fig. 7, $\gamma_{\rm cool} < \gamma_{\rm KN} < \gamma_{\rm rad,T}$, where Klein–Nishina effects feature prominently. One of these is the blue area, in which $\gamma_{\rm KN} < \gamma_{\rm rad,IC}$ and $\gamma_{\rm rad,IC} < \gamma_{\rm max} < \gamma_{\rm cool,2}$. In this area, the overall radiative cut-off Lorentz factor $\gamma_{\rm rad,IC}$ is finite (less than $\gamma_{\rm max}$) while $\gamma_{\rm cool,2}$ is not ($\gamma_{\rm cool,2} > \gamma_{\rm max}$). This means that, despite Klein–Nishina suppression of the IC cross-section, radiative losses may still regulate the highest achievable energies. To illustrate this, the contour $\gamma_{\rm rad,IC} = 4\sigma_{\rm c,0} = \gamma_{\rm X}$ is drawn (in this discussion, we assume that $\gamma_{\rm X} = 4\sigma_{\rm c,0}$ for definiteness), below which IC radiation limits direct acceleration near reconnection X-points. Straying from the blue region across the line $\gamma_{\rm max} = \gamma_{\rm rad,IC} = \gamma_{\rm cool,2}$ lands one in the red area. Here, the ordering of $\gamma_{\rm cool,2}$ and $\gamma_{\rm rad,IC}$ about $\gamma_{\rm max}$ is flipped: $\gamma_{\rm cool,2} < \gamma_{\rm max} < \gamma_{\rm rad,IC}$. Thus, the radiative cut-off energy $\gamma_{\rm rad,IC}$ is no longer finite, but the energy $\gamma_{\rm cool,2}$, beyond which particles are weakly radiative (cooling on time-scales longer than L/c), is now accessible. Here, it may be possible for high-energy particles to surpass $\gamma_{\rm cool,2}$ and enter into a weakly radiative regime. As one moves to the north-west through the red region, $\gamma_{\rm cool,2}$ becomes lower. Eventually, when

the contour $\gamma_{\rm cool,2} = 4\sigma_{\rm c,0}$ is crossed, $\gamma_{\rm cool,2}$ falls below $\gamma_{\rm X}\,(= 4\sigma_{\rm c,0})$, guaranteeing that some particles venture into the high-energy Klein–Nishina weakly radiative limit. Continuing even farther upward in the diagram, one eventually crosses the $\gamma_{\rm KN} = 2.32\gamma_{\rm cool}$ line, where the whole system becomes virtually non-radiative (case C in Fig. 7).

Finally, let us discuss Fig. 8 regions corresponding to case A of Fig. 7 ($\gamma_{\rm cool} < \gamma_{\rm rad,T} < \gamma_{\rm KN}, \gamma_{\rm max}$). The first of these is the orange area, where $\gamma_{\rm rad,IC} \simeq \gamma_{\rm rad,T} < \gamma_{\rm KN}, \gamma_{\rm X}$. Here, IC radiation limits X-point energization to below $\gamma_{\rm KN}$. In the final remaining region, the purple area, we have $4\sigma_{\rm c,0} = \gamma_{\rm X} < \gamma_{\rm rad,IC}, \gamma_{\rm KN}$. In this regime, radiative losses do not inhibit X-point acceleration, and X-point acceleration also cannot promote particles to high enough energies to stray outside the Thomson IC cooling limit.

Having overviewed the rich radiative parameter space available to IC-cooled relativistic reconnection, we now specialize to one as-yet relatively unexplored regime where pair production and Klein–Nishina radiative cooling can conspire together to form an important self-regulation mechanism. We devote the following section to a theoretical exploration of this pair-regulated Klein–Nishina radiative reconnection. We discuss applications to reconnection-driven emission from ADCe and FSRQ jets in Section 5.

4 A MODEL OF PAIR-REGULATED KLEIN–NISHINA RECONNECTION

This section explores technical aspects of reconnection with Klein–Nishina radiative cooling and pair production. The general picture is that pairs are primarily born into the upstream region, where they load the plasma energetically (i.e. the pairs are hot) but not from a number density standpoint (i.e. the pairs are tenuous). Before diving in, we state the following basic assumptions to clarify the relevant region of radiative phase space (in the sense of Fig. 8):

(i) Radiation takes place in the Klein–Nishina regime, where $\gamma_{\rm KN} < \gamma_{\rm rad,IC}$ and $\gamma_{\rm KN} < \gamma_{\rm X}$.

(ii) The reconnection region is radiatively efficient, with all particles accelerated above $\gamma_{\rm cool,1} \simeq \gamma_{\rm cool}$ cooling in less than a dynamical time L/c and most particles reaching these energies: $\langle\gamma\rangle \gg \gamma_{\rm cool}$.

(iii) The pair-production mean free paths $\lambda_{\rm mfp}$ of all gamma-rays above pair threshold are

(a) independent of photon energy, and

(b) between the full thickness Δ of *radiation zones* – the parts of the reconnection layer where above-threshold photons are produced – and the layer's full length L, i.e. $\Delta \ll \lambda_{\rm mfp} \ll L$.

As a reminder, we refer to as the 'layer' the region of the system threaded with reconnected magnetic flux.

Assumption (i) places us in the Klein–Nishina – i.e. blue or red – region of the radiative phase diagram (Fig. 8). Assumption (ii) excludes the white and green regions, implying that all of the accelerated particles – from the average energy $\langle\gamma\rangle$ to the cut-off energy – are between $\gamma_{\rm cool,1}$ and $\gamma_{\rm cool,2}$, and hence are strongly cooled. Statement (iii)(a) is not strictly true, but the pair-production cross-section $\sigma_{\gamma\gamma}(\epsilon)$ varies relatively weakly with energy beyond its peak $\sigma_{\gamma\gamma}(\epsilon) \simeq \sigma_{\rm T}/5$ when $\epsilon \simeq 3.6(m_ec^2)^2/\epsilon_{\rm ph}$. For example, $\sigma_{\gamma\gamma}(100m_e^2c^4/\epsilon_{\rm ph}) \simeq 0.08\sigma_{\rm T}$. Finally, the inequality $\lambda_{\rm mfp} \ll L$ in assumption (iii)(b) means that almost all above-threshold photons produced in the system are also absorbed in the system, and $\Delta \ll \lambda_{\rm mfp}$ further means that absorption predominantly occurs in the inflow (upstream) plasma. Note that we distinguish between the effective full thickness of the reconnection layer itself, which could be taken

as the width of the largest plasmoids $\simeq 0.1\,L$ (Uzdensky, Loureiro & Schekochihin 2010), and the thickness, Δ, of the radiation zone, which (as discussed below in Section 4.1) could be much thinner, even approaching the thicknesses of interplasmoid current layers. We illustrate the difference between Δ and $0.1L$ in Fig. 9.

In addition to all of these assumptions, we ignore effects due to synchrotron radiation. These enter at energy scales $\gamma > \gamma_s$ [see equation (28) and its surrounding discussion]. We estimate γ_s for certain astrophysical systems, and comment on the consequent limitations on the applicability of our model, in Section 5.

To investigate the basic features of reconnection in this radiative regime, we begin (Section 4.1) with some relatively simple energy-budget arguments. Based on energy considerations alone, we show that a self-regulated steady state or limit cycle should emerge – irrespective, even, of whether a pair cascade develops in the upstream region. We then decorate this basic picture by analysing the number of produced pairs. This shows (Section 4.2) that an exponential pair cascade, with each generation containing a constant factor >1 more particles than the previous one, is not expected except for (almost unrealistically) efficient particle acceleration in the reconnection layer. We further apply detailed information on the distribution of newborn pairs, showing that (Section 4.3), for $\sigma_{c,0} \ll \gamma_{KN}$, these should be fewer than those originally present in the upstream region.

4.1 The large energy density of newborn upstream pairs

Because we assume a radiatively efficient reconnection layer (ii), a sizeable fraction (e.g. one-half; Werner et al. 2019) of the inflowing Poynting flux is promptly emitted. A fraction $\mathcal{F}$ of this radiated energy lies above pair threshold with the ambient photon bath and therefore penetrates only a distance λ_{mfp} [assumption (iii)(b)] into the upstream plasma on both sides of the layer.[4] There, it is recaptured as newborn hot pairs and, ultimately, readvected into the layer. If the energy density of fresh pairs is high enough, the overall enthalpy density $w = w_0 + w_{\gamma\gamma}$ of inflowing material substantially increases. This reduces the effective hot magnetization

$$\sigma_h \equiv \frac{B_0^2}{4\pi\,w} \qquad (37)$$

below $\sigma_{h,0} = B_0^2/4\pi\,w_0$ (equation 7), which characterizes the *far upstream* region (beyond λ_{mfp} from the layer). In our convention, subscript '0' denotes far upstream quantities and subscript '$\gamma\gamma$' quantities sourced by pair creation within λ_{mfp} of the layer. Corresponding naked symbols (e.g. σ_h or w) are decided by a combination of pair-creation-sourced and far upstream values.

A reduced effective σ_h may strongly suppress the efficiency of NTPA in the layer (e.g. Guo et al. 2014, 2015; Sironi & Spitkovsky 2014; Sironi, Giannios & Petropoulou 2016; Werner et al. 2016, 2018; Werner & Uzdensky 2017; Ball, Sironi & Özel 2018). This enables a negative feedback loop, in which a layer fed initially by highly magnetized ($\sigma_{h,0} \gg 1$) plasma efficiently accelerates particles to gamma-ray emitting energies. The gamma-rays, in turn, produce pairs in the inflow region, reducing its effective magnetization and, hence, suppressing subsequent NTPA (cf. Hakobyan et al. 2019; see Fig. 1). In this section, we calculate the fixed point σ_h for this

<hr>

[4]For simplicity, we ignore kinetic beaming (Uzdensky et al. 2011; Cerutti et al. 2012b; Mehlhaff et al. 2020), which produces potentially important anisotropy in the distributions of high-energy particles and their emitted photons. We comment on expected consequences of this beaming in Appendix C but ultimately defer its full treatment to a future simulation study.
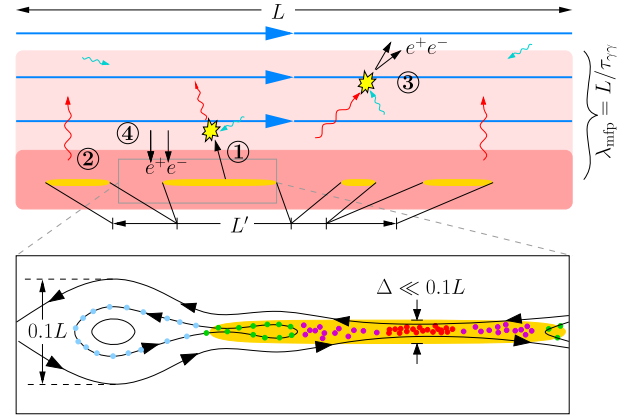
Figure 9. A more detailed view of the Klein–Nishina radiative reconnection system. The layer (region threaded with reconnected flux) contains subregions – *radiation zones* – where particles are radiating photons above pair-production threshold (gold). These occupy a combined fraction $L'/L < 1$ of the layer's full length. They are also thin, having full transverse width, $\Delta \ll 0.1L$, much smaller than the expected size of the largest plasmoids. A detailed view of a hypothetical radiation zone is displayed on the lower half of the plot. The thinness of the region ultimately stems from the kinetic-scale current sheets where particles (red dots) are accelerated near reconnection X-points. Following their impulsive acceleration in these locations, the particles (magenta dots) are magnetized by the reconnected magnetic field, which carries them away from the vicinity of an X-point. We concentrate on particles (green dots) confined to one reconnected field line. Before these particles are able to fill an entire ring in a large plasmoid where they eventually end up, they cool down below the minimum energy γ_{pp} (light blue dots) to radiate photons above pair-production threshold. The transition point where particles cool below γ_{pp} – where the colouring changes from green to light blue – determines both the length of the radiation zone (distance from the X-point) and its width, i.e. the spread in red/magenta/green particles (those with $\gamma > \gamma_{pp}$) about the mid-plane.

feedback loop. Additionally, we determine the conditions governing whether the system asymptotically approaches its fixed point in a late-time steady state. We further show that, if the fixed point is not reached, the system exhibits undamped, large-amplitude cycles of copious pair creation followed by shutdown of NTPA.

The Poynting flux delivered to the reconnection layer (per unit length in the out-of-plane direction) is

$$P_{Poynt} \sim 2L\beta_{rec}c\frac{B_0^2}{4\pi}\,. \qquad (38)$$

The leading factor of 2 results from Poynting flux entering the reconnection region from two directions. If half of this power is given to particles that quickly [within L/c; assumption (ii)] radiate it away through the IC process, the volume-averaged IC emissivity j_{IC} (power radiated per unit volume) in the reconnection layer satisfies

$$j_{IC}\Delta L' \sim \frac{1}{2}P_{Poynt} \sim L\beta_{rec}c\frac{B_0^2}{4\pi}\,, \qquad (39)$$

where L' is the combined length of all radiation zones in the reconnection layer. One can ignore all plasmoid/current-sheet substructure, taking the entire layer to be one large radiation zone, by setting $L' = L$. However, given our assumption (ii) of a radiatively efficient reconnection system, L' may actually be shorter than L. This is because particles may cool to below the minimum energy, $\gamma_{pp} \equiv 8\gamma_{KN}$ (Section 3.2.1), to emit pair-producing photons before travelling far from their primary X-point acceleration sites. Moreover, as particles travel away from an X-point, they also spread out
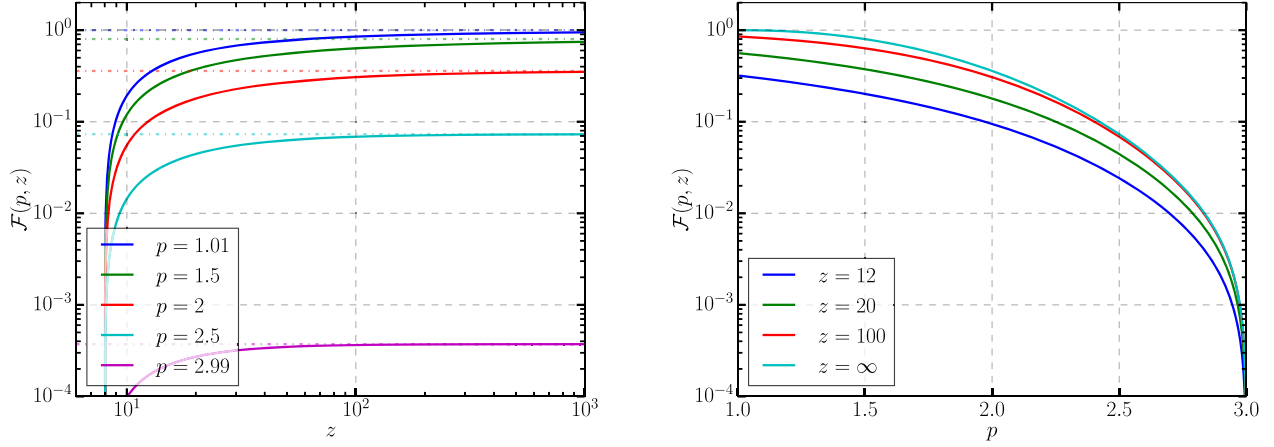
Figure 10. Left: A plot of $\mathcal{F}(p, z)$ as a function of $z \equiv \gamma_2/\gamma_{\rm KN}$ for several values of p. Dot–dashed lines indicate $\lim_{z\to\infty} \mathcal{F}(p, z)$. Right: A plot of $\mathcal{F}(p, z)$ as a function of p for several values of z. The fraction $\mathcal{F}$ is strongly dependent on p but not on z (at least after the threshold value $z = 8$ is crossed).

about the reconnection mid-plane. Thus, a cooling limit on the combined length of radiation zones (such that $L' < L$) also limits their effective thickness, Δ, potentially keeping them much thinner than the characteristic large-plasmoid width (e.g. $\Delta < 0.1L$; Fig. 9).

To determine the total enthalpy density w and, from it, the effective magnetization $\sigma_{\rm h}$ (equation 37), we need to know the fraction $\mathcal{F}$ of power radiated away from the reconnection layer above pair-production threshold (and, hence, captured in the upstream region as electron–positron pairs). Using $\gamma_{\rm pp}$ along with the distribution function of radiating layer particles $dN/d\gamma$, $\mathcal{F}$ reads

$$\mathcal{F} = \frac{\int_{\gamma_{\rm pp}}^{\infty} d\gamma \, dN/d\gamma \, P_{\rm IC}(\gamma)}{\int_1^{\infty} d\gamma \, dN/d\gamma \, P_{\rm IC}(\gamma)} . \tag{40}$$

To evaluate $\mathcal{F}$, we insert a power-law reconnection-energized pair-plasma distribution:

$$\frac{dN}{d\gamma} = A \begin{cases} \gamma^{-p} & \gamma_1 \le \gamma \le \gamma_2 \\ 0 & \text{otherwise} \end{cases} , \tag{41}$$

where A is a normalization factor and $\gamma_1 \ll \gamma_{\rm KN}$ is assumed. If $p < 3$, the γ^2-dependence of $P_{\rm IC}(\gamma)$ when $\gamma \ll \gamma_{\rm KN}$ suppresses the dependence of $\mathcal{F}$ on the onset energy $\gamma_1 \ll \gamma_{\rm KN}$ of the power law, and γ_1 can thus be taken to unity. If, instead, $p \ge 3$, the onset energy, γ_1, can also be ignored – the same dependence, $P_{\rm IC}(\gamma \ll \gamma_{\rm KN}) \propto \gamma^2$, pushes $\mathcal{F}$ to zero independently of γ_1. Thus, our assumption $\gamma_1 \ll \gamma_{\rm KN}$ is equivalent to setting $\gamma_1 = 1$.

Substituting, now, (14) and (41) into (40), as well as putting $\gamma_{\rm pp} = 8\gamma_{\rm KN}$ and $\gamma_1/\gamma_{\rm KN} = 0$, gives

$$\mathcal{F}(p, z) = \frac{\int_8^z dx \, x^{-p+2} f_{\rm KN}(x)}{\int_0^z dx \, x^{-p+2} f_{\rm KN}(x)} , \tag{42}$$

where $z \equiv \gamma_2/\gamma_{\rm KN}$. Fig. 10 displays $\mathcal{F}(p, z)$ computed according to (42). The graphs confirm the above argument that $\lim_{p\to 3} \mathcal{F}(p, z) = 0$ for all z. Furthermore, because $f_{\rm KN}(x) \sim \ln(x)/x^2$ as $x \to \infty$, when $p \le 1$, the integrals in (42) diverge with z, but in such a way that $\mathcal{F} = 1$. This signals that virtually all radiation from the layer is emitted above pair threshold. Fig. 10 also shows that, modulo a strong z-dependence near pair threshold $z = 8$, $\mathcal{F}(p, z)$ becomes nearly z-independent once $z \gtrsim 12$. Essentially, $\mathcal{F}(p, z \gtrsim 12) \simeq \mathcal{F}(p, \infty)$.

Next, we explicitly connect the fraction $\mathcal{F}$ to the effective hot magnetization $\sigma_{\rm h}$. The power $\mathcal{F} j_{\rm IC} \Delta L'$ shining out of the reconnection layer's radiation zones penetrates a distance $\lambda_{\rm mfp}$ back into the

upstream area before being deposited as hot pairs. Assuming this deposition is approximately uniform in space up to a distance $\lambda_{\rm mfp}$ above and below the reconnection layer, hot pairs add to the upstream plasma energy density at a rate $du_{\gamma\gamma}/dt$ satisfying

$$\frac{du_{\gamma\gamma}}{dt}(2\lambda_{\rm mfp})L \sim \mathcal{F} j_{\rm IC} \Delta L'$$
$$\Rightarrow \frac{du_{\gamma\gamma}}{dt} \sim \frac{\mathcal{F} j_{\rm IC}}{2} \frac{\Delta}{\lambda_{\rm mfp}} \frac{L'}{L} \sim \mathcal{F} \frac{\beta_{\rm rec} c}{\lambda_{\rm mfp}} \frac{B_0^2}{8\pi} . \tag{43}$$

In the first line, we assume that, as the radiation propagates away from the layer, it also fills in the gaps between radiation zones so that the upstream region receives pairs approximately uniformly across its length L. The second line in (43) is obtained from the first by plugging in equation (39). The factor of 2 accounts for radiated energy being absorbed both below and above the layer.

Consider a plasma parcel with initial energy density u_0 that starts far upstream, $|y| \gg \lambda_{\rm mfp}$, of the layer. The parcel is advected inward at transverse velocity $v_y = -\text{sign}(y)\beta_{\rm rec} c \simeq -\text{sign}(y)0.1c$, and, upon reaching the pair-creation zone, $|y| \sim \lambda_{\rm mfp}$, begins accruing additional energy at the rate $du_{\gamma\gamma}/dt$. The extra energy acquired in transit from $|y| \sim \lambda_{\rm mfp}$ to the layer ($|y| \sim \Delta$) is simply

$$u_{\gamma\gamma}^+ \sim \frac{1}{\beta_{\rm rec} c}(\lambda_{\rm mfp} - \Delta)\frac{du_{\gamma\gamma}}{dt} \simeq \frac{\lambda_{\rm mfp}}{\beta_{\rm rec} c} \frac{du_{\gamma\gamma}}{dt} \sim \mathcal{F} \frac{B_0^2}{8\pi} . \tag{44}$$

The accumulated internal energy density $u_{\gamma\gamma}^+$ is less than $B_0^2/8\pi$ because $\mathcal{F} \le 1$, but it can still far exceed u_0 given sufficient magnetization $\sigma_{\rm h,0} \gg 1$. The superscript '+' denotes that this is only energy *added* to the plasma; we have not yet considered that some energy may be lost en route to the layer – either through radiation or because particles physically escape the system.

Importantly, the 'readvection time' $t_{\rm ra} \equiv \lambda_{\rm mfp}/\beta_{\rm rec} c$ cancels in (44). Thus, whether the pair-creation zone is truly confined to transverse distances $|y| \sim \lambda_{\rm mfp}$ or occupies a much larger region (for example, for an N-generation pair cascade, one expects $|y| \sim \sqrt{N}\lambda_{\rm mfp} \gg \lambda_{\rm mfp}$ – a possibility that we entertain in Section 4.2), $u_{\gamma\gamma}^+$ remains approximately the same. For reference, the readvection time is related to the global dynamical time L/c through

$$t_{\rm ra} \equiv \frac{\lambda_{\rm mfp}}{\beta_{\rm rec} c} \simeq \frac{10}{\tau_{\gamma\gamma}} \frac{L}{c} , \tag{45}$$

where we used $\lambda_{\rm mfp} = L/\tau_{\gamma\gamma}$ and $\beta_{\rm rec} \simeq 0.1$. Note that the prefix 're' in 'readvection' applies only to the energy, which is captured

again by the reconnection layer. The pairs that carry this energy, by contrast, are advected into the layer for their first time.

We now estimate $u_{\gamma\gamma}$, the energy density retained by the fresh plasma swept into the reconnection layer. This yields the enthalpy density $w_{\gamma\gamma}$ and, through equation (37), the effective hot magnetization σ_h. Now, $u_{\gamma\gamma}$ is less than the deposited energy density $u_{\gamma\gamma}^+$ because, while travelling to the layer, newborn pairs may both radiatively cool and escape the system. To account for this, we define the *energy recapture efficiency*, $\xi \equiv u_{\gamma\gamma}/u_{\gamma\gamma}^+ \leq 1$, and write

$$\xi = f_{\text{nocool}} f_{\text{noesc}} . \tag{46}$$

Here, f_{nocool} and f_{noesc} are, respectively, the fraction of the accumulated energy that is not radiated away (f_{nocool}) and that is not lost through escaping particles (f_{noesc}).

We calculate the cooling factor f_{nocool} in detail in Appendix B. There, we identify a physically allowed range $f_{\text{nocool}} \in [3/400, 1]$ and show how, within this interval, f_{nocool} depends on the other parameters in the problem (on the effective magnetization σ_h and on the cut-off z). While that calculation allows us to compute σ_h self-consistently (since, in reality, f_{nocool} depends on σ_h), it is mathematically complicated. Furthermore, we find that the main qualitative features of self-regulated Klein–Nishina reconnection are captured by treating f_{nocool} as an independent parameter and scanning it across the allowed interval $[3/400, 1]$. That is the approach we adopt in this section.

In addition to this simplified prescription for f_{nocool}, we set the escape factor f_{noesc} to unity, effectively putting $\xi = f_{\text{nocool}}$. This is what one expects if the time L/c for a relativistic particle to stream out of the system is longer than the readvection time (45), which is true for $\tau_{\gamma\gamma} \gtrsim 10$ (and hence for a broad range of radiative parameters). We comment more thoroughly on the many additional kinetic effects that may influence f_{noesc} in Appendix C. However, because most of these effects tend to push f_{noesc} towards unity, we simply leave $f_{\text{noesc}} \simeq 1$ from here onward.

Using ξ, the energy density of fresh pairs entering the reconnection layer is

$$u_{\gamma\gamma} = \xi u_{\gamma\gamma}^+ \sim \xi \mathcal{F} \frac{B_0^2}{8\pi} . \tag{47}$$

If these pairs are relativistically hot, then $p_{\gamma\gamma} = u_{\gamma\gamma}/3$ and $w_{\gamma\gamma} = p_{\gamma\gamma} + u_{\gamma\gamma} \simeq (4/3)u_{\gamma\gamma}$; otherwise $w_{\gamma\gamma} = u_{\gamma\gamma}$. We take $w_{\gamma\gamma} = (4/3)u_{\gamma\gamma}$ — still a good approximation in the non-relativistic limit.

The effective inflowing plasma magnetization σ_h is then

$$\sigma_h \equiv \frac{B_0^2}{4\pi(w_0 + w_{\gamma\gamma})} = \frac{B_0^2/4\pi w_0}{1 + w_{\gamma\gamma}/w_0} \sim \frac{\sigma_{h,0}}{1 + 2\xi\mathcal{F}\sigma_{h,0}/3} . \tag{48}$$

Equation (48) encodes two main possible fixed points for σ_h. The first is when $\xi\mathcal{F}\sigma_{h,0} \ll 1$. Then, pair production is too inefficient to load the upstream plasma substantially and the solution to (48) is simply $\sigma_h \simeq \sigma_{h,0}$. The other regime is when $\xi\mathcal{F}\sigma_{h,0} \gg 1$. In this situation, hot pairs suppress σ_h to a universal value

$$\sigma_h \sim \frac{3}{2\xi\mathcal{F}} , \tag{49}$$

which is entirely independent of $\sigma_{h,0}$. Not only is (49) universal, but, in principle, it can be solved to yield self-consistent values of σ_h and p. This is because the effective magnetization governs the efficiency of NTPA (cf. Werner et al. 2016, 2018; Werner & Uzdensky 2017; Ball et al. 2018) and ultimately specifies the power-law index p. One only needs to know the reconnection NTPA 'equation of state', $p(\sigma_h)$.
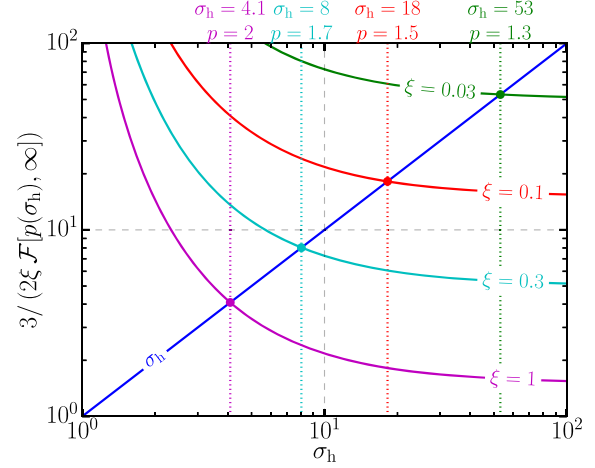

Figure 11. Solutions to equations (49) and (50) for several ξ values. Lower ξ causes the pairs born into the upstream region to cool more. This somewhat inhibits the feedback mechanism by reducing the overall enthalpy density $w = w_0 + w_{\gamma\gamma} \simeq w_{\gamma\gamma}$ of plasma arriving at the layer.

Let us assume that a suitable $p(\sigma_h)$ can be borrowed from non-radiative reconnection studies. We take

$$p(\sigma_h) = 1 + 2/\sqrt{\sigma_h} , \tag{50}$$

which can be obtained from fitting the data in fig. 3 of Werner et al. (2016) to the general form $p = C_0 + C_1/\sqrt{\sigma_h}$ used by Werner & Uzdensky (2017) (see also Ball et al. 2018; Werner et al. 2018). We acknowledge that the distribution $dN/d\gamma$ in equation (41) is the instantaneous distribution of radiating particles in the reconnection layer, which – in our radiative context – may differ from the injected (non-radiative) power-law distribution characterized by $p(\sigma_h)$. Later on, we account for approximate radiative modifications to the distribution of emitting particles. For mathematical transparency, however, in this first calculation, we plug (50) directly into our expression for $\mathcal{F}$.

To simplify further, we take $z = \gamma_2/\gamma_{\text{KN}} \to \infty$ when evaluating $\mathcal{F}$ even though calculating σ_h and p runs the same for any z. As previously remarked, the fraction $\mathcal{F}(p, z)$ is relatively z-independent as long as $z \gtrsim 12$, so taking $z \to \infty$ gives a solution representing a wide range of likely values (i.e. almost all values beyond those very close to the threshold $z = 8$ for pair production to turn on).

We now solve (49) and (50) for a variety of ξ values and graphically present the solutions in Fig. 11. A lower ξ (lower f_{nocool}) increases radiative cooling of newborn pairs as they travel towards the reconnection layer. This diminishes their enthalpy density, $w_{\gamma\gamma}$ (which, nevertheless, still dominates over the initial plasma because $w = w_0 + w_{\gamma\gamma} \simeq w_{\gamma\gamma}$), enhancing the effective magnetization σ_h, and, through $p(\sigma_h)$, hardening the resulting distribution of reconnection-energized particles.

In Fig. 11, we solve equations (49) and (50) rather than the more general form (48). This presupposes that the solution σ_h is much smaller than the original (far upstream) hot magnetization $\sigma_{h,0}$. To illustrate the effect of a finite $\sigma_{h,0}$, we also display solutions to equation (48) for $\sigma_{h,0} = 100$ in Fig. 12. As expected, a finite $\sigma_{h,0}$ has relatively little impact on the value of σ_h when $\sigma_h \ll \sigma_{h,0}$ – the universal regime in which the solution is insensitive to the far upstream magnetization. However, as the resulting solution σ_h gets closer to $\sigma_{h,0}$, the approximate solution obtained from (49) becomes less accurate. This occurs roughly when $\xi \sim 1/\sigma_{h,0}$ (compare, for example, the solutions obtained for $\xi = 0.03 = 3/\sigma_{h,0}$ in Figs 11
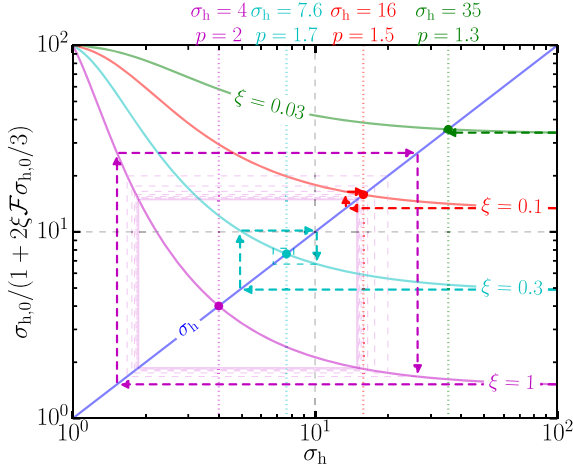
Figure 12. Solutions to the more general equations (48) and (50) using $\sigma_{h,0}$ = 100. Unlike equation (49), equation (48) does not assume *a priori* that $\sigma_h \ll \sigma_{h,0}$. The figure also depicts the stability of each solution. Systems starting with magnetization $\sigma_{h,0}$ transition to a new magnetization $\sigma_{h,1} = h(\sigma_{h,0})$ after approximately one readvection time $\lambda_{mfp}/\beta_{rec}c$. This is illustrated for each ξ by a horizontal line running left from $h(\sigma_{h,0})$ to the corresponding new value of $\sigma_{h,1}$ on the blue diagonal [which represents the curve $x = h(x)$]. Then, after another readvection time, the second modified magnetization $\sigma_{h,2} = h(\sigma_{h,1})$ is reached. This is represented by both an upward-running dashed line from $\sigma_{h,1}$ to $h(\sigma_{h,1})$ and a horizontal line, from left to right, terminating on the corresponding value $\sigma_{h,2}$ on the diagonal. Further transitions are denoted by one vertical and one horizontal dashed line: either down-then-left or up-then-right. For low ξ, the system approaches the fixed point σ_h after just a few transitions. For high ξ, pair feedback is so efficient that the system gets stuck in a two-state swing cycle.

and 12). In addition, Fig. 12 illustrates the stability of the fixed point σ_h, which is the topic of the next section.

4.1.1 Stability of the pair-loaded σ_h

Now that we know how to calculate the fixed point σ_h, we can also begin to ask whether a system that starts from the initial magnetization $\sigma_{h,0}$ actually approaches σ_h at some late time. We call the fixed point 'stable' if the plasma feeding the reconnection layer approaches a quasi-steady magnetization σ_h, and 'unstable' otherwise. Where necessary, we further distinguish 'global stability', which refers to the notion of stability just described, from 'local stability', which is simpler, and only determines whether a system that starts at a magnetization some infinitesimal distance $\delta\sigma_h$ away from σ_h approaches σ_h.

To discuss stability quantitatively, we abbreviate the right-hand side of (48) as

$$h(x) \equiv \frac{\sigma_{h,0}}{1 + 2\xi\,\mathcal{F}[p(x),\infty]\,\sigma_{h,0}/3}. \tag{51}$$

A system that begins with initial magnetization $\sigma_{h,0}$ (i.e. before any pairs have been produced) accelerates a distribution of particles in the reconnection layer with a power-law index $p(\sigma_{h,0})$ given by equation (50). These particles then radiate photons, some of which are above pair threshold with the background radiation and consequently – after a time $\sim\lambda_{mfp}/c$ – pair produce somewhere in the zone $|y| \lesssim \lambda_{mfp}$. The newborn pairs are then advected towards the layer, reaching it after a readvection time $t_{ra} = \lambda_{mfp}/\beta_{rec}c$. At that point the layer witnesses a new effective magnetization $\sigma_{h,1}$ that is determined by $\sigma_{h,0}$ via $\sigma_{h,1} = h(\sigma_{h,0})$. Assuming that the time for the layer

to respond to a new magnetization is smaller than $\lambda_{mfp}/\beta_{rec}c$, the lag time between when the layer starts processing a $\sigma_{h,0}$-plasma and when it starts to see a $\sigma_{h,1}$-plasma is $\lambda_{mfp}/c + \lambda_{mfp}/\beta_{rec}c \simeq \lambda_{mfp}/\beta_{rec}c$. Thus, the readvection time characterizes the transition period from the initial magnetization $\sigma_{h,0}$ to the first modified magnetization $\sigma_{h,1}$.

By similar reasoning, after about $n\lambda_{mfp}/\beta_{rec}c$, the layer witnesses effective magnetization

$$\sigma_{h,n} = h(\sigma_{h,n-1}). \tag{52}$$

Thus, the system approaches the fixed point σ_h as an asymptotic steady state if $\lim_{n\to\infty}\sigma_{h,n} = \sigma_h$. Of course, if the convergence is slow, then the system may not actually reach the fixed point σ_h (even though it is stable) before reconnection finishes. All that is certain in that case is that oscillatory swings about σ_h are damped: with each successive readvection time, the system reaches a magnetization that is somewhat closer to the asymptotic steady state. However, if the long-time limit of the iterated map $\sigma_{h,n} = h(\sigma_{h,n-1})$ does not approach the fixed point σ_h, then some other behaviour occurs. That late-time outcome is large-amplitude, undamped oscillations about σ_h between two states, one with a low magnetization $\sigma_{h,<} < \sigma_h$ and one with a high value $\sigma_{h,>} > \sigma_h$. The high-$\sigma_{h,>}$ state drives efficient NTPA in the reconnection layer and, consequently, pair production in the upstream region. When the created hot pairs reach the layer, they initiate the low-$\sigma_{h,<}$ state, where NTPA is shut down, pair creation ceases, and, eventually (after another $\lambda_{mfp}/\beta_{rec}c$), the high-$\sigma_{h,>}$ state is restored. The cycle repeats from there.

Both behaviours – an asymptotic steady state and a late-time two-state cycle – are illustrated in Fig. 12. For low-energy recapture efficiency ξ, the system tends towards a steady state; when ξ is increased beyond a critical threshold, the system bifurcates, asymptotically favouring large-amplitude swings. This can be understood as follows. For efficient feedback, so much of the energy radiated from the layer is caught in, and then recaptured from, the upstream region that the system overshoots the fixed point σ_h by a wide margin. This chokes subsequent pair-producing emission from the layer effectively enough that the system violently ricochets back to a high magnetization, getting caught between two extreme magnetizations $\sigma_{h,<}$ and $\sigma_{h,>}$ as described above.

We now show how this physical picture is encoded in our mathematical machinery. As is well known from the theory of iterated maps, the fixed point $\sigma_h = h(\sigma_h)$ is locally stable if $|h'(\sigma_h)| \leq 1$ (meaning, as discussed before, that magnetizations $\sigma_h + \delta\sigma_h$ move towards σ_h with each successive iteration). Higher ξ lowers the curve $h(x)$, which increases the magnitude of the slope $h'(\sigma_h)$ (i.e. makes it more negative) at the location of the fixed point (see Fig. 12), pushing the system towards the instability threshold $|h'(\sigma_h)| = 1$ for swing cycles. Even though this is only a local stability condition, we argue in Appendix D that, for the restricted class of monotonically decreasing $h(x)$ functions, local instability is sufficient for global instability: σ_h is never approached by systems that start at $\sigma_{h,0}$ when $|h'(\sigma_h)| > 1$. In that case, the late-time dynamics consist of a two-state swing cycle. In contrast, if $|h'(\sigma_h)| < 1$, the system may or may not converge towards σ_h depending on the details of the map function $h(x)$. However (as also argued in Appendix D), at late times, all monotonically decreasing $h(x)$ functions result in either a steady state or a swing cycle. For such functions, *no other late-time dynamics are allowed*.

To illustrate these remarks, we supply Fig. 13. That figure displays the fixed point σ_h as a function of ξ. The system bifurcates at a critical value of $\xi = \xi_c \simeq 0.84$: the fixed point σ_h becomes unstable [$|h'(\sigma_h)|$ goes above 1] and a new attractor for the late-time dynamics appears – a two-state cycle characterized by a low magnetization $\sigma_{h,<} < \sigma_h$ and
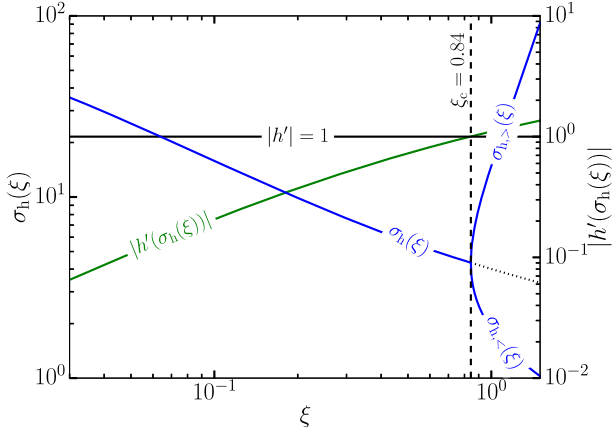
Figure 13. The fixed point σ_h of the reconnection layer [i.e. of the iterated map $\sigma_{h,n} = h(\sigma_{h,n-1})$ with $h(x)$ as defined in equation 51] plotted as a function of ξ. Also displayed are the magnitude of the slope of $h(x)$ evaluated at the fixed point and the low-$\sigma_{h,<}$ and high-$\sigma_{h,>}$ magnetizations associated with two-state cycles. At the critical value $\xi = \xi_c \simeq 0.84$, the system undergoes a bifurcation: the fixed point σ_h goes unstable, as signalled by $|h'(\sigma_h(\xi))|$ crossing above 1. The fixed point still exists when $\xi > \xi_c$ (and is indicated by a dotted line in that case), but does not control the late-time dynamics. Instead, the system comes to hop between two magnetizations $\sigma_{h,<} < \sigma_h$ and $\sigma_{h,>} > \sigma_h$.

a high magnetization $\sigma_{h,>} > \sigma_h$. Importantly, the condition $|h'| > 1$ implies that the system traps into a limit cycle at late times. While, in general, it is *not* the case that $|h'| < 1$ implies convergence to the fixed point σ_h, for this particular system that happens to be true. (We discuss a case where this does not hold in Appendix B.)

4.1.2 Calculating σ_h with more realistic radiative feedback

Having established the mathematical techniques for computing the self-regulated magnetization σ_h and determining its stability, we now repeat our calculation with more physical realism. In particular, we modify our assumed layer particle distribution $dN/d\gamma$ from a single power law, as in (41), to a broken power law

$$\frac{dN}{d\gamma} = A \begin{cases} (\gamma/\gamma_{KN})^{-p_T} & \gamma \leq \gamma_{KN} \\ (\gamma/\gamma_{KN})^{-p_{KN}} & \gamma > \gamma_{KN} \end{cases}. \tag{53}$$

This allows us to account for the different expected modifications to the power-law scaling of the distribution function in both the Thomson ($\gamma \leq \gamma_{KN}$) and deep Klein–Nishina ($\gamma > \gamma_{KN}$) regimes.

Given a steady injected particle distribution with power-law index p, the realized distribution, if cooled in the Thomson limit, is steepened to index $p_T = p + 1$. In contrast, if one adopts the approximate form for $f_{KN}(q)$ from equation (17), then, for $\gamma > \gamma_{KN}$, the actual distribution of radiating particles, in fact, hardens, attaining a power-law scaling $p_{KN} \simeq p - 0.5$ (Moderski et al. 2005). We therefore take $p_T = p(\sigma_h) + 1$ and $p_{KN} = p(\sigma_h) - 0.5$ where $p(\sigma_h) = 1 + 2/\sqrt{\sigma_h}$ as in equation (50). Admittedly, this is still very crude. It ignores the finite time-scale of particle energization in the reconnection layer and the finite time it takes for the full energy range of the distribution function to respond to radiative losses. It ignores, moreover, the bursty nature of magnetic reconnection at the highest energies (e.g. Werner et al. 2019; Mehlhaff et al. 2020). However, these relationships for p_T and p_{KN} are at least a first step towards capturing some of the qualitative effects that radiative cooling may have on the pair feedback mechanism.

Using (53), the power fraction becomes (cf. equation 42)

$$\tilde{\mathcal{F}}(p_T, p_{KN}, z) \equiv \frac{\int_8^z dx\, x^{-p_{KN}+2} f_{KN}(x)}{\int_0^1 dx\, x^{-p_T+2} f_{KN}(x) + \int_1^z dx\, x^{-p_{KN}+2} f_{KN}(x)}. \tag{54}$$

Note that $\tilde{\mathcal{F}}(p_T, p_{KN}, z \leq 8) = 0$, and so it is not necessary to consider the case $z < 1$ in the denominator. The fraction $\tilde{\mathcal{F}}$, with $p_T = p + 1$ and $p_{KN} = p - 0.5$ is displayed as a function of z and p in Fig. 14. The opposite impacts of radiative losses in the Thomson and deep Klein–Nishina limits – where, respectively, the particle energy distribution tends to steepen and become shallower – have a pronounced effect on the p-dependence of $\tilde{\mathcal{F}}$. Due to the steepening of the distribution at lower energies, the denominator diverges when $p > 2$ ($p_T > 3$). Meanwhile, due to the hardening in the Klein–Nishina regime, the fraction $\tilde{\mathcal{F}}$ goes to 1 as $z \to \infty$ when $p < 1.5$ ($p_{KN} < 1$). This compresses the effective range in p over which $\tilde{\mathcal{F}}$ varies. Whereas our original power fraction $\mathcal{F}(p, \infty)$ in equation (42) starts to decline from 1 only once $p \gtrsim 1.5$, not reaching zero until $p = 3$, our new fraction $\tilde{\mathcal{F}}(p + 1, p - 0.5, \infty)$ departs from 1 when $p > 1.5$ but hits zero already by the time $p = 2$.

This dramatically impacts the fixed points $\sigma_h(\xi)$ and their stabilities. In analogy to equation (51), we define

$$\tilde{h}(x) = \frac{\sigma_{h,0}}{1 + 2\xi\, \tilde{\mathcal{F}}[p(x) + 1, p(x) - 0.5, \infty]\sigma_{h,0}/3}. \tag{55}$$

Fig. 15 displays a few $\tilde{h}(x)$ curves, each one with different energy recapture efficiency ξ. As promised, $\tilde{h}(x)$ is flat when $p(x) = 1 + 2/\sqrt{x} > 1.5$ (i.e. when $x > 16$) because it is here that $\tilde{\mathcal{F}} = 1$. At the same time, $\tilde{h}(x < 4) = \sigma_{h,0}$ since $p(x < 4) > 2$, and, hence, $\tilde{\mathcal{F}} = 0$. The rapid transition in $\tilde{h}(x)$ from nearly vertical to nearly flat between $x = 4$ and $x = 16$ induces a sharp transition in the stability of the fixed point σ_h. The fixed point becomes unstable at $\xi = \xi_c \simeq 0.30$, and further increasing ξ beyond this point yields much more dramatic swing cycles than did incremental changes in ξ beyond the threshold $\xi_c \simeq 0.84$ of h in equation (51). By the time $\xi \gtrsim 0.5$, the reconnection layer *starts* in a swing cycle, hopping between its *initial* magnetization $\sigma_{h,0} = \sigma_{h,>}$ and a much lower magnetization $\sigma_{h,<} < 4$. In the low-magnetization state, pair production completely ceases ($\tilde{\mathcal{F}} = 0$) and, after one readvection time, the plasma flowing into the layer contains no newborn pair component and once again possesses the initial magnetization $\sigma_{h,0}$.

Although we are still a long way from a detailed model, in this section, we have uncovered what appears to be a robust mechanism for pair feedback on Klein–Nishina relativistic reconnection. An initially highly magnetized system $\sigma_{h,0} \gg 1$ can efficiently accelerate gamma-ray-radiating leptons in the reconnection layer. These gamma-rays collide with ambient background photons to produce a hot newborn pair component in the upstream plasma. Subsequently, the new pairs are advected into the layer where they suppress NTPA and, hence, the production of additional pairs. This mechanism operates even when radiative cooling of the layer particles is taken into account. If the energy recapture efficiency ξ approaches order unity, then the system undergoes undamped, large-amplitude oscillations between a high magnetization (even as high as the initial value $\sigma_{h,0}$) – sourcing copious upstream pairs – and a low magnetization where pair production is completely shut down. For $\xi \ll 1$ however, the system approaches a steady state – characterized by a balance between upstream pair-loading and particle energization in the layer – after just one or a few readvection times $t_{ra} = \lambda_{mfp}/\beta_{rec}c$.
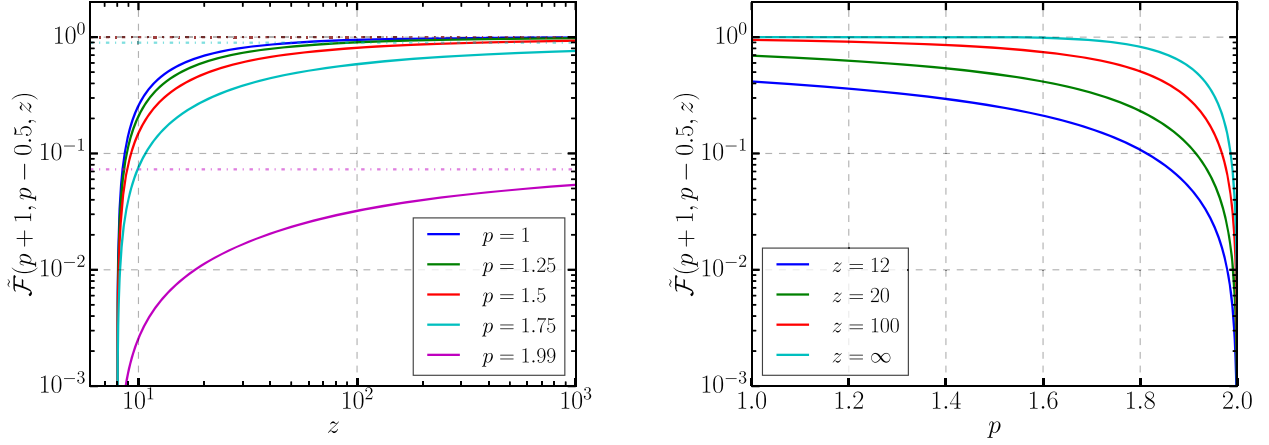
Figure 14. Left: A plot of $\tilde{\mathcal{F}}(p+1, p-0.5, z)$ as a function of $z \equiv \gamma_2/\gamma_{\mathrm{KN}}$ for several p. Dot–dashed lines indicate $\lim_{z\to\infty} \tilde{\mathcal{F}}(p+1, p-0.5, z)$. Right: A plot of $\tilde{\mathcal{F}}(p, z)$ as a function of p for several z. Due to the opposite influences of radiative cooling in the Thomson and deep Klein–Nishina regimes, where the particle energy distribution tends to be softened and hardened, respectively, the range in p over which $\tilde{\mathcal{F}}$ significantly varies compresses relative to that of $\mathcal{F}$ ($1.5 < p < 2$ for the former instead of $1.5 < p < 3$ for the latter).
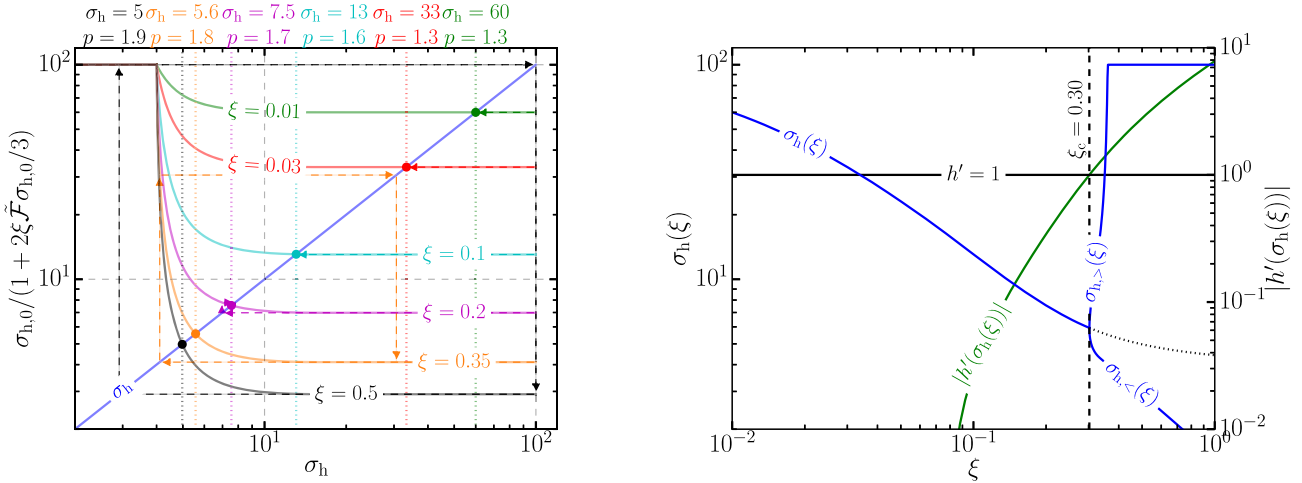

Figure 15. Left: The analogue of Fig. 12 for $\tilde{\mathcal{F}}$ (as defined in equation 54) instead of $\mathcal{F}$ (as defined in equation 42). Right: The analogue of Fig. 13 for $\tilde{\mathcal{F}}$. The tendency for radiative losses to steepen the particle distribution in the Thomson regime and, in contrast, to harden it in the deep Klein–Nishina limit creates a sharp and dramatic bifurcation. The critical value of the energy recapture efficiency $\xi_c \simeq 0.30$ is smaller than $\xi_c = 0.84$ obtained in Fig. 13, indicating that radiation back reaction makes the system more 'touchy': more susceptible to late-time limit cycles. Once ξ exceeds this threshold, the system rapidly transitions to a two-state cycle featuring very large amplitude swings between the initial magnetization $\sigma_{\mathrm{h},>} = \sigma_{\mathrm{h},0}$ and a small magnetization $\sigma_{\mathrm{h},<} \sim 1$.

4.2 Upstream pair cascades are not generally expected

So far, we have described a fixed point solution in which the reconnection layer regulates itself, maintaining a universal effective hot magnetization σ_{h}, or, potentially, exhibiting large-amplitude two-state limit cycles. The arguments that allowed us to characterize this behaviour were based solely upon tracing the flow of energy through the system. We now undertake an analogous program, tracing the plasma particles instead of the energy.

Here, we must deal with an additional complication that was not present in the preceding energy-based arguments. Namely, because of pair production, particle number is not conserved, and one cannot simply equate the number of pairs born into the inflow plasma with the number of above-threshold photons radiated away from the layer. In principle, additional pair creation in the upstream region can occur: newborn pairs can radiate additional photons that are themselves above pair-production threshold and capable of producing secondary pair generations. We address this issue in this

section. As an immediate byproduct of the analysis, we show that a pair cascade is not generally expected – the number of particles only grows exponentially in each subsequent pair generation under certain optimal conditions.

An important quantity here is the distribution of particles (electrons + positrons) injected in the inflow plasma as the result of high-energy photons, originally emitted from the reconnection layer, getting absorbed in the upstream region. We denote this as

$$Q_{\gamma\gamma}^{(1)}(\gamma) = B_1 \begin{cases} \gamma^{-\Gamma_1} & \gamma_1^{(1)} \le \gamma < \gamma_2^{(1)} \\ 0 & \text{otherwise} \end{cases}. \tag{56}$$

Here, the superscript '(1)' (subscript '1' on Γ_1 and B_1) indicates the first generation. In this section, we calculate

$$Q_{\gamma\gamma}^{(n)}(\gamma) = B_n \begin{cases} \gamma^{-\Gamma_n} & \gamma_1^{(n)} \le \gamma < \gamma_2^{(n)} \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

the injected distribution of nth-generation pairs, in terms of $Q_{\gamma\gamma}^{(1)}(\gamma)$. The dimensions of $Q_{\gamma\gamma}^{(n)}(\gamma)$ are that of a rate: particles per unit time per unit energy. We have not normalized by spatial volume. Thus, $Q_{\gamma\gamma}^{(n)}(\gamma)$ is averaged over the spatial region where nth-generation pair production is active. For now, we assume that $Q_{\gamma\gamma}^{(n)}(\gamma)$ is a power law and check this a posteriori.

4.2.1 Newborn pair generations: basic observations

Before calculating the $Q_{\gamma\gamma}^{(n)}(\gamma)$ distributions in detail, we offer a few guiding remarks. These primarily pertain to the range of energies present in each generation – i.e. to $\gamma_1^{(n)}$ and $\gamma_2^{(n)}$ – and to the number of generations that the upstream region can support.

As discussed in Section 3, a particle with Lorentz factor $\gamma \geq \gamma_{\rm pp} = 8\gamma_{\rm KN}$ emits photons of characteristic energy $E_\gamma \sim \gamma m_e c^2/2$. Upon absorption by the soft radiation background, these photons each create two particles with approximate Lorentz factors $E_\gamma/2m_e c^2 \sim \gamma/4$. Thus, the power law (56) begins and ends at energies about a factor of 4 less than the corresponding energies of layer particles: $\gamma_1^{(1)} \sim \gamma_{\rm pp}/4 = 2\gamma_{\rm KN}$ and $\gamma_2^{(1)} \sim \gamma_2/4$, ($\gamma_2$ is the cut-off layer particle energy; see equation 41).

In general, the low and high injected energies in each successive pair generation follow by identical reasoning to those of the first generation. One has $\gamma_1^{(n)} \sim \gamma_1^{(1)} \sim \gamma_{\rm pp}/4 = 2\gamma_{\rm KN}$ and $\gamma_2^{(n)} \sim \gamma_2/4^n$. The last generation, N, possesses cut-off at or below the threshold to emit pair-producing radiation: $\gamma_{\rm pp} = 8\gamma_{\rm KN} \geq \gamma_2^{(N)} = \gamma_2/4^N$. This caps the number of generations to $N = \lceil \log_4(\gamma_2/8\gamma_{\rm KN}) \rceil$, where $\lceil x \rceil$ rounds x up to the nearest integer. Note that, if $\log_4(\gamma_2/8\gamma_{\rm KN})$ is not an integer, then a partial additional generation is produced from the subset of pairs in the preceding generation with Lorentz factors above $\gamma_{\rm pp}$. For example, if $\log_4(\gamma_2/8\gamma_{\rm KN})$ is 2.4 – and, hence, $N = 3$ – then two full generations are created in addition to a partial third generation.

One may also place a potentially firmer limit on the number of generations. Note that newborn upstream particles with $t_{\rm cool,IC}(\gamma) > t_{\rm ra}$ do not radiate before entering the layer, and, thus, do not yield additional upstream pairs. This condition is marginally satisfied if $\gamma = \tilde{\gamma}_{\rm cool}$ such that

$$t_{\rm cool,IC}(\tilde{\gamma}_{\rm cool}) \equiv \frac{10}{\tau_{\gamma\gamma}} \frac{L}{c} \simeq \frac{\lambda_{\rm mfp}}{\beta_{\rm rec} c} , \tag{58}$$

where, for convenience, we use the approximate form $t_{\rm ra} \simeq 10L/\tau_{\gamma\gamma}c$ (equation 45). Note that the cooling time $t_{\rm cool,IC}$ is the product of the readvection time with a dimensionless function of $q = \gamma/\gamma_{\rm KN}$: namely, $t_{\rm cool,IC}(\gamma) = (3/50)(10L/\tau_{\gamma\gamma}c)[q f_{\rm KN}(q)]^{-1}$, which follows from equations (31) and (45). This implies that the solution to (58) depends only on $\gamma_{\rm KN}$ – it is independent of all other system parameters (e.g. $\gamma_{\rm max}$, $\gamma_{\rm rad,T}$, and $\sigma_{\rm c,0}$). Moreover, because $t_{\rm cool,IC}(\gamma)$ is non-monotonic (see equation 31, Fig. 5, and the surrounding discussion), there are actually two solutions to (58). These can be found numerically and are

$$\tilde{\gamma}_{\rm cool,1} \simeq 0.066\gamma_{\rm KN} \simeq (3/50)\gamma_{\rm KN} , \tag{59}$$

which is in the Thomson limit, and

$$\tilde{\gamma}_{\rm cool,2} \simeq 280\gamma_{\rm KN} , \tag{60}$$

which is in the deep Klein–Nishina regime.

All pairs born into the upstream region with $\gamma > \tilde{\gamma}_{\rm cool,2}$ have $t_{\rm cool,IC}(\gamma) > t_{\rm ra}$ and thus do not radiate before being swept into the reconnection layer. However, pairs born with $\gamma < \tilde{\gamma}_{\rm cool,2}$ (which also have $\gamma > \tilde{\gamma}_{\rm cool,1}$ because $\gamma \geq \gamma_1^{(n)} \sim 2\gamma_{\rm KN} > \tilde{\gamma}_{\rm cool,1}$) have $t_{\rm cool,IC}(\gamma) < t_{\rm ra}$. They maintain this condition as they radiatively

cool all the way until they reach $\gamma = \tilde{\gamma}_{\rm cool,1}$, at which point $t_{\rm cool,IC}(\gamma) = t_{\rm ra}$. Thus, pairs born between the energies $\tilde{\gamma}_{\rm cool,1}$ and $\tilde{\gamma}_{\rm cool,2}$ may cool all the way down to $\tilde{\gamma}_{\rm cool,1}$ before entering the layer. (They may not cool quite this far if not born a full mean-free-path $\lambda_{\rm mfp}$ upstream of the layer.)

Because first-generation pairs whose parent layer particles had Lorentz factors $\geq 4\tilde{\gamma}_{\rm cool,2}$ do not spawn additional pairs, the number of generations is limited to $N = \lceil \log_4(4\tilde{\gamma}_{\rm cool,2}/8\gamma_{\rm KN}) \rceil \simeq \lceil \log_4(282/2) \rceil = 4$. Combining this with the previous limit yields

$$N = \min \left[\left\lceil \log_4 \left(\frac{\gamma_2}{8\gamma_{\rm KN}} \right) \right\rceil , 4 \right] , \tag{61}$$

which slightly modifies the cut-off energies from $\gamma_2^{(n)} \sim \gamma_2/4^n$ to

$$\gamma_2^{(n)} \sim \begin{cases} \gamma_2/4 & n = 1 \\ \min(\gamma_2, 4\tilde{\gamma}_{\rm cool,2})/4^n & n > 1 \end{cases} . \tag{62}$$

While the first generation's cut-off is determined entirely in terms of the layer cut-off γ_2, only particles in the first generation with $\gamma < \tilde{\gamma}_{\rm cool,2}$ can give rise to further generations before entering the layer.

The maximum number of generations N places an important constraint on the self-regulated reconnection system. The mean excursion of nth-generation photons (those that create nth-generation pairs) from the reconnection layer is $|\bar{y}^{(n)}| \sim \sqrt{n}\lambda_{\rm mfp}$, which scales with n as a 1D random walk. Thus, photons would begin to escape the system if $N \geq (L/\lambda_{\rm mfp})^2 \gg 1$. But, because N cannot exceed 4, this is not expected.

4.2.2 Newborn pair generations: detailed calculation

We now calculate, in detail, the $Q_{\gamma\gamma}^{(n)}(\gamma)$ distributions in terms of $Q_{\gamma\gamma}^{(1)}(\gamma)$. To do this, we set up a system of coupled differential equations for $N_{\gamma\gamma}^{(n)}(\gamma)$, the (volume-integrated) distributions of pairs born into the inflow plasma, and for $N_{\rm ph}^{(n)}(\epsilon)$, the (volume-integrated) distributions of photons residing in the upstream region. To formulate these equations, we temporarily add time-dependence to the distributions though we ultimately specialize to the steady state. The time-derivatives of the $N_{\gamma\gamma}^{(n)}$'s can be expressed as

$$\frac{\partial}{\partial t} N_{\gamma\gamma}^{(n)}(\gamma, t) = (\text{injection}) - (\text{cooling})$$

$$= Q_{\gamma\gamma}^{(n)}(\gamma, t) - \frac{\partial}{\partial \gamma} \left(\dot{\gamma} N_{\gamma\gamma}^{(n)}(\gamma, t) \right) . \tag{63}$$

Here, we have fictitiously assumed that the discrete Klein–Nishina-limit cooling of particles can be represented continuously. However, this approximation has proven to be quite accurate for non-mono-energetic pair distributions $N_{\gamma\gamma}^{(n)}$ (Zdziarski 1989; Moderski et al. 2005). The signed cooling rate is given by $-\dot{\gamma} = |\dot{\gamma}| = \gamma/t_{\rm cool,IC}(\gamma)$.

Simultaneously, the time-derivatives of the $N_{\rm ph}^{(n)}$'s are

$$\frac{\partial}{\partial t} N_{\rm ph}^{(n)}(\epsilon, t) = (\text{emission}) - (\text{annihilation})$$

$$= 2\frac{N_{\gamma\gamma}^{(n-1)}(2\epsilon, t)}{t_{\rm cool,IC}(2\epsilon)} - \frac{N_{\rm ph}^{(n)}(\epsilon, t)}{\lambda_{\rm mfp}/c} , \tag{64}$$

where, in this section only, we write photon energies ϵ in units of $m_e c^2$. The leading factor 2 on the first term comes from the Jacobian $d\gamma/d\epsilon = 2$ corresponding to $\gamma = 2\epsilon$. Equation (64) is written explicitly in the Klein–Nishina regime, where each upstream pair cools in time $t_{\rm cool,IC}(\gamma)$, and, when it does, emits a photon of energy $\gamma m_e c^2/2$. Then, each photon travels a distance $\lambda_{\rm mfp}$ in time $\lambda_{\rm mfp}/c$ before annihilating against the background to produce

a new pair. Because 2 particles are injected into the nth generation upon each annihilation of an nth-generation photon, one has

$$Q_{\gamma\gamma}^{(n)}(\gamma, t) = 2 \times 2 \frac{N_{\text{ph}}^{(n)}(2\gamma, t)}{\lambda_{\text{mfp}}/c} . \tag{65}$$

Here again, one factor of 2 comes from the Jacobian from particle to photon energies.

Equations (63)–(65) achieve a steady state if the readvection time, over which the $n = 1$ injection term is constant, is longer than the cooling times of newborn pairs and the photon streaming time λ_{mfp}/c. The first condition is satisfied because all pairs (at least for $n > 1$) are born with Lorentz factors $\gamma < \tilde{\gamma}_{\text{cool},2}$, while the latter is satisfied because the nth-generation readvection time is $\sqrt{n}\lambda_{\text{mfp}}/\beta_{\text{rec}}c \sim 10\sqrt{n}\lambda_{\text{mfp}}/c \gg \lambda_{\text{mfp}}/c$. (The particles in the $n = 1$ generation with $\gamma > \tilde{\gamma}_{\text{cool},2}$ do not produce additional pairs and so can be excluded from these equations.) Let us therefore specialize to the steady state of equations (63) and (64). This gives, upon plugging (65) into (64),

$$Q_{\gamma\gamma}^{(n)}(\gamma) = 8 \frac{N_{\gamma\gamma}^{(n-1)}(4\gamma)}{t_{\text{cool,IC}}(4\gamma)} , \tag{66}$$

which, in turn, allows one to write

$$\frac{\partial}{\partial \gamma} \left(\dot{\gamma} N_{\gamma\gamma}^{(n)}(\gamma) \right) = -\frac{1}{8} \frac{\partial}{\partial \gamma} \left(\gamma Q_{\gamma\gamma}^{(n+1)}(\gamma/4) \right) , \tag{67}$$

where we have used $-\dot{\gamma} = \gamma/t_{\text{cool,IC}}(\gamma)$.

Then, inserting (67) into the steady state of (63) gives a recursive formula for the $Q_{\gamma\gamma}^{(n)}(\gamma)$'s:

$$-\frac{1}{8} \frac{\partial}{\partial \gamma} \left(\gamma Q_{\gamma\gamma}^{(n+1)}(\gamma/4) \right) = Q_{\gamma\gamma}^{(n)}(\gamma). \tag{68}$$

Let us integrate this equation from $\gamma \geq \gamma_{\text{pp}}$ to $\gamma = \infty$. The upper bound gives zero on the left-hand side of (68) since $Q_{\gamma\gamma}^{(n)}(\infty) = 0$. The lower bound does *not* give zero on the left because $\gamma/4$ exceeds the lowest energy $\gamma_1^{(n)} = 2\gamma_{\text{KN}}$ in the $(n + 1)$st injected distribution. Hence,

$$Q_{\gamma\gamma}^{(n+1)}(\gamma) = \frac{2}{\gamma} \int_{4\gamma}^{\infty} Q_{\gamma\gamma}^{(n)}(\gamma')\mathrm{d}\gamma' . \tag{69}$$

Let us now use (69) to determine $Q_{\gamma\gamma}^{(n+1)}(\gamma)$ explicitly, assuming $Q_{\gamma\gamma}^{(n)}(\gamma)$ is given by a power law as in equation (57). Restricting to $2\gamma_{\text{KN}} = \gamma_1^{(n+1)} \leq \gamma < \gamma_2^{(n+1)} = \gamma_2^{(n)}/4$, we have

$$Q_{\gamma\gamma}^{(n+1)}(\gamma) = \frac{2B_n}{\gamma} \int_{4\gamma}^{\gamma_2^{(n)}} (\gamma')^{-\Gamma_n}\mathrm{d}\gamma'$$

$$= \frac{2B_n}{\gamma(\Gamma_n - 1)}(4\gamma)^{-\Gamma_n + 1} \left[1 - \left(\frac{4\gamma}{\gamma_2^{(n)}} \right)^{\Gamma_n - 1} \right] . \tag{70}$$

For γ not between $2\gamma_{\text{KN}}$ and $\gamma_2^{(n)}/4$, the distribution $Q_{\gamma\gamma}^{(n)}(\gamma)$ is zero. Strictly speaking, for $n = 1$, $\gamma_2^{(1)} = \gamma_2/4$ needs to be modified to $\min(\gamma_2^{(1)}, \tilde{\gamma}_{\text{cool},2})/4$, since only first-generation particles with energies less than $\tilde{\gamma}_{\text{cool},2}$ can spawn additional pairs.

Noting that the term in square brackets in (70) is roughly unity except when $\gamma \sim \gamma_2^{(n)}/4$, we see that $Q_{\gamma\gamma}^{(n+1)}(\gamma)$ is just a power law with a cut-off at $\gamma_2^{(n+1)} = \gamma_2^{(n)}/4$, in agreement with equation (62). This verifies the assumed power-law form (57) provided the initial injected distribution $Q_{\gamma\gamma}^{(1)}(\gamma)$ is also a power law. In that case,

$$\Gamma_{n+1} = \Gamma_n = \Gamma_1 \equiv \Gamma \tag{71}$$

and

$$B_{n+1} = \mathcal{A}(\Gamma)B_n = [\mathcal{A}(\Gamma)]^n B_1 , \tag{72}$$

where we have defined

$$\mathcal{A}(\Gamma) \equiv \frac{8}{4^{\Gamma}(\Gamma - 1)} . \tag{73}$$

These recurrence formulae imply that the total pair injection rate in the upstream region is

$$\sum_{n=1}^{N} \int Q_{\gamma\gamma}^{(n)}(\gamma)\mathrm{d}\gamma = B_1 \sum_{n=1}^{N} [\mathcal{A}(\Gamma)]^{n-1} \frac{\gamma_{\text{min}}^{-\Gamma+1}}{\Gamma - 1} \left[1 - \left(\frac{\gamma_{\text{min}}}{\gamma_2^{(n)}} \right)^{\Gamma-1} \right]$$

$$\simeq \frac{B_1 \gamma_{\text{min}}^{-\Gamma+1}}{\Gamma - 1} \sum_{n=0}^{N-1} [\mathcal{A}(\Gamma)]^n$$

$$\simeq \left[\int Q_{\gamma\gamma}^{(1)}(\gamma)\mathrm{d}\gamma \right] \sum_{n=0}^{N-1} [\mathcal{A}(\Gamma)]^n$$

$$= \frac{1 - [\mathcal{A}(\Gamma)]^N}{1 - \mathcal{A}(\Gamma)} \int Q_{\gamma\gamma}^{(1)}(\gamma)\mathrm{d}\gamma , \tag{74}$$

where, for convenience, we define $\gamma_{\text{min}} \equiv \gamma_1^{(n)} = 2\gamma_{\text{KN}}$. In the second and third lines, we assume that $\Gamma > 1$ and, hence, that the term $(\gamma_{\text{min}}/\gamma_2^{(n)})^{\Gamma-1}$ can be neglected. The final line is the crux of this section. It determines under what conditions a true pair cascade develops – when the total number of injected pairs in the upstream region is exponential in the number of generations N. We have,

$$\mathcal{A}(\Gamma) > 1 \iff \text{pair cascade} ,$$
$$\mathcal{A}(\Gamma) < 1 \iff \text{no pair cascade} . \tag{75}$$

The multiplication factor $\mathcal{A}(\Gamma)$ is less than unity for $\Gamma > 1.73$. Whether a pair cascade develops comes down to the expected value of Γ in the first-generation injected distribution.

We expect that $Q_{\gamma\gamma}^{(1)}(\gamma)$ inherits its power-law scaling from the distribution of photons emitted into the upstream region from the layer. For a power-law distribution (53) of radiating layer particles, the emitted photon spectrum is, approximately, a power law with index $\Gamma = p_{\text{KN}} + 1$ (plus a logarithmic correction; see Blumenthal & Gould 1970; Aharonian & Atoyan 1981). Since we expect the scaling of the layer particle distribution in the Klein–Nishina regime to be $p_{\text{KN}} = p - 0.5$, which corresponds to an intrinsic particle acceleration index p hardened due to Klein–Nishina IC losses, we predict $\min(\Gamma) = \min(p + 0.5)$. Now, according to (50), $\min(p) = 1$, and thus the distribution of first-generation injected pairs should be no harder than $\min(\Gamma) = 1.5$. This gives a maximum multiplication factor of $\max(\mathcal{A}(\Gamma)) = \mathcal{A}(\min(\Gamma)) = \mathcal{A}(1.5) = 2$, for which the number of injected pairs doubles in each successive generation.

A shallow enough scaling Γ to bring $\mathcal{A}(\Gamma)$ above unity is only achieved for $p < 1.7 - 0.5 = 1.2$. Using equation (50), this implies a magnetization $\sigma_{\text{h}} > 80$. Also, the theory we have presented thus far ignores the possibility of a guide field, which tends to suppress reconnection-powered NTPA (e.g. Werner & Uzdensky 2017). Thus, a pair cascade is possible, but requires quite optimal combinations of parameters (high-σ_{h} and small guide field, for example).

4.3 The small number density of upstream pairs

In this section, we employ our tally of the upstream pair-creation rate – given by the $Q_{\gamma\gamma}^{(n)}(\gamma)$ distributions – to answer another important question associated with Klein–Nishina radiative reconnection.

Namely, we calculate the pair-production *multiplicity*

$$\eta \equiv \frac{n_{\gamma\gamma}}{n_0}, \tag{76}$$

the ratio of the number density $n_{\gamma\gamma}$ of newborn pairs entering the reconnection layer to that n_0 of pairs in the far upstream region.

Let us define the volumetric (per unit time per unit volume) rate of pair production $dn_{\gamma\gamma}^{(n)}/dt$ into the nth generation. Assuming that nth-generation pairs are deposited uniformly up to a transverse distance $\bar{y}^{(n)}$ (e.g. $\bar{y}^{(n)} \sim \sqrt{n}\lambda_{\mathrm{mfp}}$) away from the layer across the full width L of the system, we have

$$\frac{dn_{\gamma\gamma}^{(n)}}{dt} \sim \frac{1}{2\bar{y}^{(n)}L} \int d\gamma \, Q_{\gamma\gamma}^{(n)}(\gamma). \tag{77}$$

Noting that a parcel of plasma travels from $|y| \sim \bar{y}^{(n)}$ to the layer at $|y| \sim \Delta \ll \bar{y}^{(n)}$ over time $\bar{y}^{(n)}/\beta_{\mathrm{rec}}c$, during which it accrues nth-generation pairs at the rate $dn_{\gamma\gamma}^{(n)}/dt$, we have

$$n_{\gamma\gamma} \sim f_{\mathrm{noesc}} \sum_{n=1}^{N} \frac{\bar{y}^{(n)}}{\beta_{\mathrm{rec}}c} \frac{dn_{\gamma\gamma}^{(n)}}{dt} \sim \frac{1}{2\beta_{\mathrm{rec}}cL} \int d\gamma \sum_{n=1}^{N} Q_{\gamma\gamma}^{(n)}(\gamma). \tag{78}$$

In the second step, we assume $f_{\mathrm{noesc}} = 1$. Similar to the cancellation of the readvection time in (44), $\bar{y}^{(n)}$ cancels in (78).

To evaluate (78), we normalize the $Q_{\gamma\gamma}^{(n)}(\gamma)$'s by balancing the particle-creation rate in the *first* generation with twice the number of above-threshold photons emitted from the reconnection layer:

$$\int d\gamma \, Q_{\gamma\gamma}^{(1)}(\gamma) = 2\Delta L' \int_{\gamma_{\mathrm{pp}}} d\gamma \, R_{\mathrm{IC}}(\gamma)dN/d\gamma, \tag{79}$$

where $dN/d\gamma$ is the distribution of radiating particles as in equations (41) and (53). In turn, one may normalize the distribution $dN/d\gamma$ (to find A in equation 41 or 53) by setting the power radiated from the layer to $P_{\mathrm{Poynt}}/2$ (as in equations 38 and 39):

$$\frac{1}{2}P_{\mathrm{Poynt}} = L\beta_{\mathrm{rec}}c\frac{B_0^2}{4\pi} \sim j_{\mathrm{IC}}\Delta L'$$

$$= \Delta L' \int d\gamma \, P_{\mathrm{IC}}(\gamma)dN/d\gamma. \tag{80}$$

In keeping with the earlier parts of this study, we consider two cases: one in which the radiation reaction force on layer particles is ignored and one in which it is approximately included. In the former case, we assume a single power-law form for $dN/d\gamma$, as in equation (41), with index $p(\sigma_{\mathrm{h}}) = 1 + 2/\sqrt{\sigma_{\mathrm{h}}}$ (equation 50). In the latter, we adopt a broken power law, as in equation (53), with separate indices $p_{\mathrm{T}} = p(\sigma_{\mathrm{h}}) + 1$ and $p_{\mathrm{KN}} = p(\sigma_{\mathrm{h}}) - 0.5$ in the Thomson ($\gamma < \gamma_{\mathrm{KN}}$) and Klein–Nishina ($\gamma > \gamma_{\mathrm{KN}}$) regimes, respectively. To save space, we simultaneously conduct both analyses by plugging in equation (53) for $dN/d\gamma$ and leaving p_{T} and p_{KN} unspecified until the end of the calculation.

Proceeding in this manner, we find that

$$n_{\gamma\gamma} \sim n_0 \frac{\sigma_{\mathrm{c},0}}{\gamma_{\mathrm{KN}}} \tilde{\mathcal{M}}(p_{\mathrm{T}}, p_{\mathrm{KN}}, z), \tag{81}$$

where $z = \gamma_2/\gamma_{\mathrm{KN}}$. The right-hand side does not depend on n_0: the n_0 in the numerator cancels against that in the definition of $\sigma_{\mathrm{c},0}$. The *multiplicity function* $\tilde{\mathcal{M}}(p_{\mathrm{T}}, p_{\mathrm{KN}}, z)$ is derived, as discussed, by evaluating (78) whilst enforcing (79) and (80). It reads

$$\tilde{\mathcal{M}}(p_{\mathrm{T}}, p_{\mathrm{KN}}, z) \equiv \frac{3\int_8^z dx \, g_{\mathrm{KN}}(x)x^{-p_{\mathrm{KN}}}}{\int_0^1 dx \, f_{\mathrm{KN}}(x)x^{-p_{\mathrm{T}}+2} + \int_1^z dx \, f_{\mathrm{KN}}(x)x^{-p_{\mathrm{KN}}+2}}$$

$$\times \frac{\int d\gamma \sum_{n=1}^{N} Q_{\gamma\gamma}^{(n)}(\gamma)}{\int d\gamma \, Q_{\gamma\gamma}^{(1)}(\gamma)}. \tag{82}$$

The factor on the second line encodes the possibility of a pair cascade and is roughly equal to $[1 - \mathcal{A}(\Gamma)^N]/[1 - \mathcal{A}(\Gamma)]$, as in equation (74). However, here we evaluate this factor by maintaining finite z-dependent cut-offs in each term, writing

$$\int d\gamma \, Q_{\gamma\gamma}^{(n)}(\gamma) \simeq \frac{[\mathcal{A}(\Gamma)]^{n-1} B_1 \gamma_{\mathrm{min}}^{-\Gamma+1}}{\Gamma - 1} \left[1 - \left(\frac{\gamma_{\mathrm{min}}}{\gamma_2^{(n)}}\right)^{\Gamma-1}\right]. \tag{83}$$

We do not simplify the right-hand side to $\mathcal{A}(\Gamma)^{n-1} B_1 \gamma_{\mathrm{min}}^{-\Gamma+1}/(\Gamma - 1)$, as done in Section 4.2 (equation 74). This ensures that contributions from each generation turn on gradually (as they do in reality), thereby keeping the multiplicity function continuous.

To denote the case where we ignore radiative feedback on the layer particles, in which we put $p_{\mathrm{T}} = p_{\mathrm{KN}} = p$, we write

$$\mathcal{M}(p, z) \equiv \tilde{\mathcal{M}}(p, p, z). \tag{84}$$

We display the functions $\mathcal{M}(p, z)$ and $\tilde{\mathcal{M}}(p + 1, p - 0.5, z)$ in Fig. 16. Similar to the case of $\mathcal{F}$ and $\tilde{\mathcal{F}}$, the separate impacts of radiative cooling in the Thomson and deep Klein–Nishina limits contract the range in p across which the multiplicity function varies. Just as for the power fraction $\mathcal{F}$, the steepening of the layer particle distribution $p_{\mathrm{T}} = p + 1$ in the Thomson regime completely shuts down pair production whenever $p_{\mathrm{T}} > 3$ (i.e. $p > 2$). However, unlike the power fraction $\mathcal{F}$, the multiplicity function exhibits a much more complicated non-monotonic dependence on the underlying parameters p and z. In addition to this, while the power fraction attains order unity for a wide range of p and z, the multiplicity function never does – it is never larger than $\simeq 0.1$ for any parameter combination, and is very often much smaller than this.

This last fact means that, even when a pair cascade truly does develop in the upstream region – and the total newborn pair count is exponential in the number of generations – the overall multiplicity may still be *small*. In fact, because the multiplicity function obtains a global maximum of order 10^{-1}, we can map out precisely the parameters for which η is guaranteed to be small: $\gamma_{\mathrm{KN}} > \sigma_{\mathrm{c},0}/10$. As we show in Section 5, this condition is roughly satisfied for reconnection in both FSRQ jets and black hole ADCe.

This concludes our detailed discussion of pair-regulated Klein–Nishina magnetic reconnection. Next, we examine the potential astrophysical ramifications of reconnection in this regime.

5 ASTROPHYSICAL IMPLICATIONS

In this section, we discuss observational aspects of pair-regulated Klein–Nishina magnetic reconnection. Our approach has two stages. First, we elaborate observable consequences of our model that are generic, not requiring an explicit global astrophysical context. This grounds our subsequent discussion where we estimate parameters for concrete astrophysical systems – FSRQs and black hole ADCe – and comment on observations that the model may help to explain.

5.1 General observable features

Here, we discuss the appearance of Klein–Nishina radiative reconnection as viewed through a telescope. We assume that the reconnection system is not spatially resolved.

5.1.1 Observed radiation comes mostly from the layer

In radiative reconnection without pair-production feedback, the radiative output is dominated by high-energy particles in the layer (i.e.
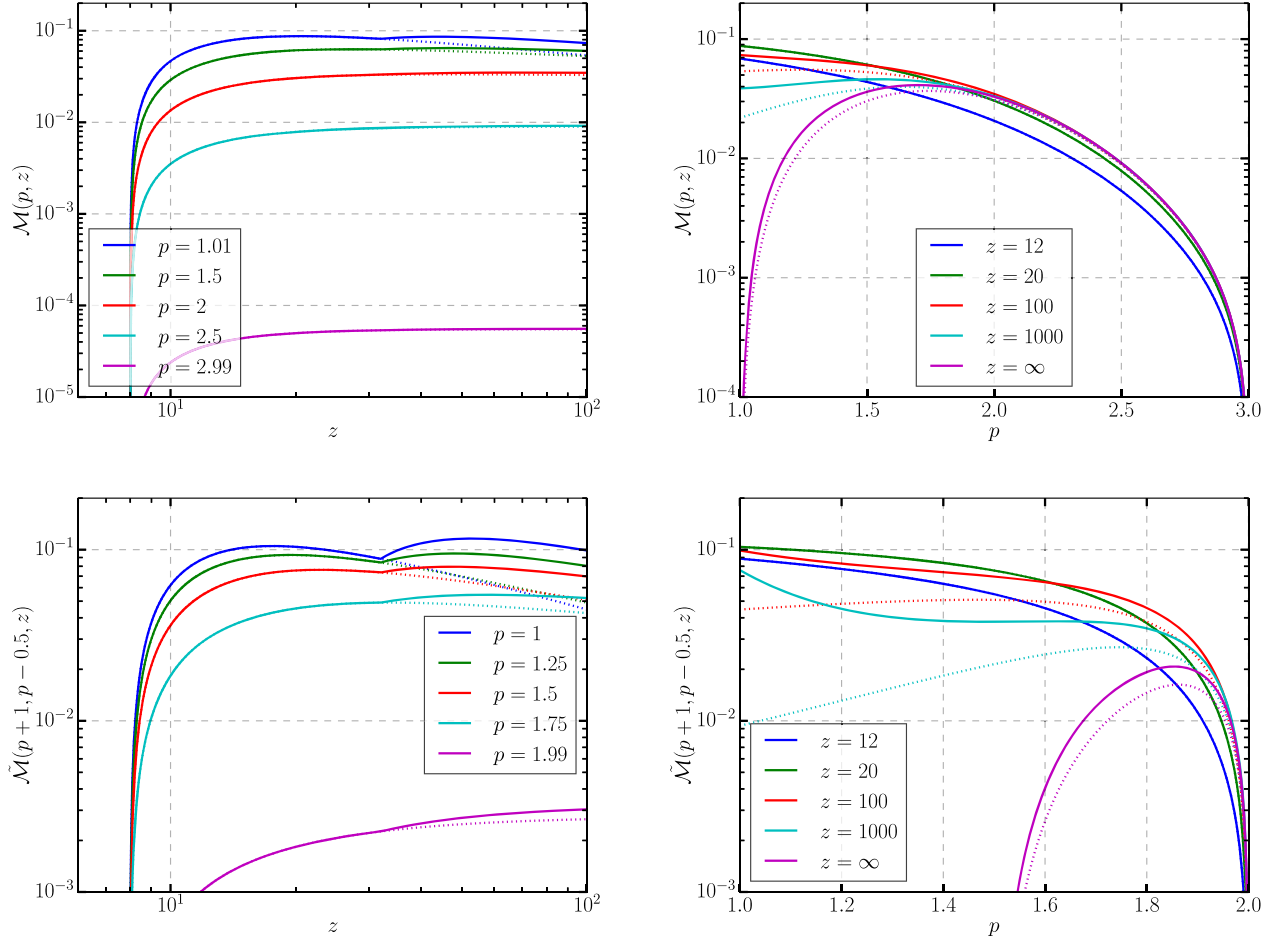
Figure 16. The analogue of Figs 10 and 14, which respectively display $\mathcal{F}$ and $\tilde{\mathcal{F}}$, for $\mathcal{M}$ and $\tilde{\mathcal{M}}$. In all plots, dotted lines show the value of the multiplicity function divided by the pair cascade factor on the second line of equation (82), demonstrating what the multiplicity function would be if only accounting for pairs injected into the first upstream generation (i.e. ignoring a possible pair cascade). Top left: $\mathcal{M}(p, z) \equiv \tilde{\mathcal{M}}(p, p, z)$ displayed as a function of z for several p. Top right: $\mathcal{M}(p, z)$ displayed as a function of p for several z. Bottom left: $\tilde{\mathcal{M}}(p + 1, p - 0.5, z)$ displayed as a function of z for several p. Bottom right: $\tilde{\mathcal{M}}(p + 1, p - 0.5, z)$ displayed as a function of p for several z. While the dependence of $\mathcal{M}$ and $\tilde{\mathcal{M}}$ on p and z is more complicated (in particular, non-monotonic) than that of $\mathcal{F}$ and $\tilde{\mathcal{F}}$, the former are always small – no larger than 0.1 – and therefore the multiplicity η is small whenever $\gamma_{KN} > \sigma_{c,0}/10$. Pair cascades significantly influence the multiplicity for large z and small p_{KN} and, hence, have a more pronounced impact on $\tilde{\mathcal{M}}$, where $p_{KN} = p - 0.5$, than on $\mathcal{M}$, where $p_{KN} = p$.

the downstream region permeated by reconnected magnetic flux). However, with pair feedback, the upstream and downstream plasmas are radiatively coupled: a substantial portion of the energy emitted from the layer may be intercepted upstream and reprocessed into high-energy newborn pairs. While en route to the layer, these pairs emit potentially observable light (if $f_{nocool} < 1$) that could, in principle, outshine the escaping (below-threshold) radiation from the layer.

Therefore, we wish to determine whether the light that an observer sees comes predominantly from the upstream or downstream (layer) plasmas. We denote the respective *observable* luminosities (i.e. only of below-threshold and, hence, escaping radiation) of these two regions per unit length in the out-of-plane direction by $L_{upstream}$ and L_{layer}. The first luminosity is the fraction of the layer's above-threshold radiated power that is captured by the upstream region and reemitted below threshold before flowing back into the layer:

$$L_{upstream} \sim (1 - f_{nocool})(2L\lambda_{mfp})\frac{du_{\gamma\gamma}}{dt}$$

$$\sim (1 - f_{nocool})\mathcal{F}\beta_{rec}cL\frac{B_0^2}{4\pi} \tag{85}$$

(cf. equation 43). Meanwhile, the fraction of the layer's power emitted below pair threshold is (cf. equation 39)

$$L_{layer} \sim (1 - \mathcal{F})\frac{1}{2}P_{Poynt} \sim (1 - \mathcal{F})\beta_{rec}cL\frac{B_0^2}{4\pi}. \tag{86}$$

Note that one may substitute $\tilde{\mathcal{F}}$ for $\mathcal{F}$ in these expressions provided the same symbol is used in both. The luminosity ratio is

$$\frac{L_{upstream}}{L_{layer}} \sim (1 - f_{nocool})\frac{\mathcal{F}}{1 - \mathcal{F}}. \tag{87}$$

For the upstream region to outshine the layer, both significant above-threshold radiation ($\mathcal{F} > 1/2$) and relatively low-energy recapture efficiency ($f_{nocool} \ll 1$) are required.

However, it is also possible that the system enters a two-state cycle, swinging between a low and a high effective upstream magnetization (see Section 4.1). Then, the upstream domain may appear brightest in one magnetization state while the downstream region shines the most in the other state. In this case, the relevant luminosities to compare are probably the brightest luminosities achieved by each region (even if in opposite states). This is certainly appropriate if the

readvection time t_{ra} is not resolved by the observations. However, even if the readvection time is resolved, different locations along the layer's surface (e.g. different areas in the horizontal direction of Fig. 9) may be decorrelated from each other. One radiating zone may be in a high-magnetization phase while a neighbouring zone is in the corresponding low-magnetization state. Then, the observed luminosities from each zone are averaged and (assuming, for simplicity, a 50 per cent duty cycle) dominated by the brighter state of the more luminous region (upstream or downstream) in that zone.

Thus, the luminosity ratio in equation (87) may not accurately describe a reconnection layer prone to swing cycles. In fact, according to equation (86), the layer is observationally brightest in its *low*-magnetization state – this is when $\mathcal{F}$ is smallest and, hence, when most of the incident Poynting flux is radiated at low enough energies (below pair threshold) to escape the system. In contrast, the upstream region becomes brightest when it receives an abundant supply of above-threshold photons – when the layer is in its *high* magnetization phase. These are processed into an energetic fresh pair plasma component that subsequently radiates below threshold [unless $f_{nocool}(\sigma_{h,>}) \simeq 1$]. Thus, we expect that, in a two-state swing cycle, the appropriate luminosity ratio is

$$\frac{L_{\text{upstream}}(\sigma_{h,>})}{L_{\text{layer}}(\sigma_{h,<})} \sim [1 - f_{\text{nocool}}(\sigma_{h,>})] \frac{\mathcal{F}(\sigma_{h,>})}{1 - \mathcal{F}(\sigma_{h,<})}$$
$$\sim [1 - f_{\text{nocool}}(\sigma_{h,>})] \leq 1, \qquad (88)$$

where, in the last step, we loosely approximated $\mathcal{F}(\sigma_{h,>}) \sim (1 - \mathcal{F}(\sigma_{h,<})) \sim 1$, which is often roughly correct in swing cycles (see Section 4). In the above, we also consider f_{nocool} to depend on the magnetization. The detailed dependence (discussed in Appendix B) is not important here beyond that, generally, $f_{nocool}(\sigma_{h,<}) \ll 1$ and $f_{nocool}(\sigma_{h,>}) \sim 1$. Thus, in a swing cycle, the layer probably dominates the observed radiation.

An exceptional case occurs if the swing cycle high state has energy recapture efficiency $\xi(\sigma_{h,>}) \simeq f_{nocool}(\sigma_{h,>})$ close enough to unity to render the upstream region brightest in the low-magnetization state {i.e. if $[1 - f_{nocool}(\sigma_{h,>})]\mathcal{F}(\sigma_{h,>}) < [1 - f_{nocool}(\sigma_{h,<})]\mathcal{F}(\sigma_{h,<})$}. Then the relevant luminosity ratio, instead of equation (88), is $L_{\text{upstream}}(\sigma_{h,<})/L_{\text{layer}}(\sigma_{h,<})$, which evaluates to $[1 - f_{nocool}(\sigma_{h,<})]\mathcal{F}(\sigma_{h,<})/[1 - \mathcal{F}(\sigma_{h,<})] \sim \mathcal{F}(\sigma_{h,<}) \ll 1$, and is quite dominated by the layer.

Thus, the layer generally produces most of the observable radiation from pair-regulated Klein–Nishina reconnection. The upstream luminosity only dominates in a steady state if $\mathcal{F} > 1/2$ and $f_{nocool} \ll 1$, and it may at most be comparable to the layer luminosity in an asymptotic swing cycle. This seems to be corroborated by observations because the expected spectrum produced from the upstream region has spectral index $\alpha = 1/2$ (calculated in Appendices B and E), but the objects we discuss below have steeper scalings (as the layer might produce in a low-magnetization state).

5.1.2 Klein–Nishina physics may promote rapid variability through kinetic beaming

In our earlier work (Mehlhaff et al. 2020), we performed a detailed study of the interplay between radiative physics (Thomson IC radiation reaction) in reconnection and the kinetic beaming phenomenon first discovered by Cerutti et al. (2012b). As detailed in both references, in collisionless relativistic reconnection, the electromagnetic fields near reconnection X-points tend to simultaneously accelerate *and* collimate particles. The higher energy particles are focused more tightly than the lower energy particles, making this beaming

inherently kinetic. A collimated beam of high-energy particles may then sweep across an observer's line of sight, and the synchrotron or IC emission of the bunch – also emitted as a beam – may then create a dramatic blip in the measured light curve: a 'lighthouse effect'.

In Mehlhaff et al. (2020), we showed that efficient radiative losses play a critical role in enabling the kinetic beaming mechanism to impact observations. Without strong radiative cooling, collimated beams of particles isotropize before dumping their reconnection-acquired energy into energetic photons, and thus most of their radiation is emitted quasi-isotropically. Only efficiently cooled beamed particle bunches can radiate their energy before dispersing, leading to sweeping beams of light that may manifest as rapid flares.

We here recall two specific criteria from our earlier work that may be necessary for, or at least promote, observable signatures of kinetic beaming:

(i) Particles may need to have cooling times within a certain multiple of their gyroperiods in the reconnecting magnetic field B_0. (If this multiple is 1, then this is the *saturated cooling* condition from Section 2.2.2 – particles are at their radiation-limited energy $\gamma_{rad,IC}$.) The number of gyroperiods before a collimated bunch of particles isotropizes may be large but still finite. In fact, one main result from Mehlhaff et al. (2020) is that kinetic beaming only manifests in the emission from particles with energies within about a decade of the radiative saturation Lorentz factor $\gamma_{rad,T}$, corresponding to $t_{cool,T} \lesssim 100$ gyroperiods.

(ii) Direct acceleration (by the reconnection electric field) near reconnection X-points must deliver particles to higher energies than secondary acceleration channels (in the language of Sections 2 and 3, $\gamma_{sec} \ll \gamma_{rad,IC}$, γ_X – secondary channels must radiatively stall before both the X-point radiative and intrinsic acceleration limits). Otherwise, secondary – and presumably more isotropic – energization processes may wash out signatures of kinetic beaming in the reconnection-energized distribution of particles.

Our present work provides at least three reasons why pair-regulated Klein–Nishina reconnection could meet these criteria, perhaps leading to even more pronounced kinetic beaming than when particles are subject to purely Thomson radiative cooling.

(1) The first reason pertains to point (i) above. Namely, Klein–Nishina effects render a particle's *cooling ratio* – the ratio of its cooling length to its Larmor radius – very insensitive, when $\gamma \gg \gamma_{KN}$, to its Lorentz factor. To wit,

$$\lim_{\gamma \gg \gamma_{KN}} \frac{ct_{cool,IC}(\gamma)}{2\pi\gamma\rho_0} = \lim_{\gamma \gg \gamma_{KN}} \left(\frac{\gamma_{rad,T}}{\gamma}\right)^2 \frac{1}{f_{KN}(\gamma/\gamma_{KN})}$$
$$= \frac{10}{2\pi} \frac{2}{9} \frac{(\gamma_{rad,T}/\gamma_{KN})^2}{\ln(\gamma/\gamma_{KN}) - 11/6}$$
$$\simeq 0.4 \frac{(\gamma_{rad,T}/\gamma_{KN})^2}{\ln(\gamma/\gamma_{KN}) - 11/6}, \qquad (89)$$

where we used approximation (16). Even when Lorentz factors as high as $\gamma_{rad,IC}$, for which $ct_{cool,IC}/2\pi\gamma\rho_0 = 10/2\pi \sim 1$, are not accessible, particles may still reach a high-energy ($\gamma \gg \gamma_{KN}$) regime where their cooling times come within some moderately large, beaming-favouring multiple of their gyroperiods, and where their cooling ratios become essentially γ-independent. In such a scenario, a broad range (e.g. from γ_{KN} to γ_X) of energetic particles may radiate efficiently enough to be kinetically beamed. This range could possibly be broader than in the Thomson regime, for which $t_{cool,T}(\gamma)$ declines as γ^{-1}, making it increasingly difficult to accelerate particles up to higher and higher energies. However, one should invoke

this argument with some caution because the smallest accessible cooling ratio in the Klein–Nishina regime may be quite large, scaling, according to (89), as $(\gamma_{\rm rad,T}/\gamma_{\rm KN})^2$.

(2) A second way Klein–Nishina effects may promote kinetic beaming pertains to item (ii) above. As argued in Section 3.2, Klein–Nishina radiative cooling selectively suppresses secondary, slower acceleration channels relative to rapid, impulsive acceleration near reconnection X-points. In the Klein–Nishina regime, the IC cooling time $t_{\rm cool,IC}(\gamma)$ grows as $\gamma/\ln(\gamma)$ (equation 32), which effectively removes the radiative cap on X-point acceleration (because $t_{\rm X}$ is also proportional to γ; see equation 4). In contrast, secondary energization mechanisms generally operate on time-scales that grow more quickly with γ (e.g. as γ^2), and so maintain a finite cut-off, even for deeply Klein–Nishina radiative cooling.

(3) Finally, and also relevant to point (ii) above, the energy distribution of produced pairs is very broad and non-thermal (Section 4.2). These particles serve precisely as the pre-accelerated upstream population that, as argued by Mehlhaff et al. (2020), may help overcome the conventional $\gamma_{\rm X} \simeq 4\sigma_{\rm c}$ limit. Upon entering the reconnection layer, newborn particles possess Larmor radii ($\geq \tilde{\gamma}_{\rm cool,1}\rho_0$) that may exceed those, $\sim\sigma_{\rm c}\rho_0$, of typical accelerated particles – most of which come from the much colder and more numerous particle population (Section 4.3) that was already present in the far upstream region. If so, then newborn pairs sample larger field structures than the elementary current layers and plasmoids, with size scale $\sigma_{\rm c}\rho_0$, at the bottom of the plasmoid hierarchy (Werner et al. 2016; Uzdensky 2020). Unlike the vast majority of initially present cold upstream particles, these pairs may surf across many elementary layers – that together comprise a much larger acceleration region – becoming energized well beyond $4\sigma_{\rm c}$ before finally becoming magnetized. Thus, $\gamma_{\rm X}$ is effectively raised. This is important because higher $\gamma_{\rm X}$ enables direct acceleration to energize and collimate higher energy particles, potentially increasing the energy range of kinetically beamed particles and photons.

Detailed predictions for kinetic beaming are outside the scope of this paper; our analysis only traces the largest scale bulk flow of energy and particles through the reconnection system. However, these simple observations provide a target for future dedicated simulations, which may study, in detail, the effects of Klein–Nishina and pair-production physics on kinetic beaming.

5.1.3 Caution against detailed spectral predictions

While we believe that the basic qualitative features of our model presented in Section 4 are fairly robust, the model is not presently quantitatively accurate enough to warrant making specific spectral predictions for various astrophysical sources. The specific power-law scalings of the particle and photon spectra depend on a number of uncertain details.

For one thing, the NTPA 'equation of state' $p(\sigma_{\rm h}) = 1 + 2/\sqrt{\sigma_{\rm h}}$ from equation (50) is quite crude and only intended for illustration. The steady-state power-law scalings $p_{\rm T} = p + 1$ and (especially) $p_{\rm KN} = p - 0.5$ are also just rough estimates. For example, the bursty nature of reconnection is already known to modify $p_{\rm T}$ when Thomson IC cooling is quite strong (Werner et al. 2019).

As a related issue, the power-law index of the photon spectrum radiated by particles above $\gamma_{\rm KN}$ (in the $p_{\rm KN}$ portion of the particle distribution) is only very loosely given by $\Gamma = p_{\rm KN} + 1 = p + 0.5$. In fact, Moderski et al. (2005) estimate a harder power-law $\Gamma = p$ and predict little change between the power-law scaling of the emitted photon spectrum from the Thomson to Klein–Nishina regimes. Thus,

given the level of quantitative uncertainty in our model, one should avoid inferring numerical values of plasma parameters (like $\sigma_{\rm h}$ and $z = \gamma_2/\gamma_{\rm KN}$) from observed photon spectra, especially at the highest energies (above $\gamma_{\rm KN}m_ec^2$). It is also probably not warranted to test the model by searching for a prominent Klein–Nishina spectral break.

Having discussed now some generic observational features of our model as well as some of its quantitative limitations (with respect to interpreting power-law scalings of photon spectra), we now turn to some concrete classes of astrophysical systems.

5.2 Consequences for specific astrophysical systems

In this section, we explicitly estimate reconnection energy scales ($\gamma_{\rm max}$, $\sigma_{\rm c,0}$, $\gamma_{\rm rad,T}$, $\gamma_{\rm KN}$, etc.) for two types of astrophysical systems – FSRQ jets and black hole ADCe. We also comment on the prospects for our model either to explain certain observational phenomena or to help constrain astrophysical details that are difficult to pin down from observations alone.

5.2.1 The radiative environments of FSRQs

FSRQs form a subclass of blazars, which are active galactic nuclei that launch counterpropagating relativistic jets, one of which is directed towards the observer. Blazar spectra generally feature a characteristic double-humped structure (Fossati et al. 1998; Ghisellini 2011). The low-energy component is usually understood as synchrotron radiation by electrons and positrons in the jet. The high-energy peak is also commonly associated with jet leptons but via a different process: IC scattering of soft seed photons (Begelman, Fabian & Rees 2008; Böttcher et al. 2013; Madejski & Sikora 2016). Leptonic models differ on whether the particles producing both spectral components are cospatial and on the source of Compton seed photons. In BL Lacs, lower luminosity blazars with low- and high-energy humps peaked at relatively high frequencies (UV/X-ray and hundreds of GeV gamma-rays, respectively; Ghisellini 2011; Madejski & Sikora 2016), the seed photons are generally thought to be furnished by synchrotron emission within the jet itself (synchrotron self-Compton models; e.g. Maraschi, Ghisellini & Celotti 1992; Bloom & Marscher 1996). For the higher luminosity sources with lower energy spectral peaks (often falling in the IR at low energies and a few hundred MeV in the gamma-rays; Ghisellini 2011; Madejski & Sikora 2016), the FSRQs, typical models envision Compton seed photons impinging from the circumnuclear environment (Begelman & Sikora 1987; Melia & König 1989; Sikora, Begelman & Rees 1994).

We focus on FSRQs. In addition to their higher luminosities and lower energy spectra (relative to BL Lacs), FSRQs possess strong emission lines attributed to an accretion disc. The disc illuminates the jet directly and also shines on to clouds of circumnuclear material – the broad emission line region (also 'broad-line region'; BLR) and dusty torus (also 'hot dust region'; HDR). These structures reprocess the accretion disc light and redirect some of its energy back on to the jet, thereby providing strong sources of external radiation situated relatively far from the central engine (e.g. Nalewajko, Begelman & Sikora 2014; Madejski & Sikora 2016).

The smaller of the two circumnuclear regions is the BLR, which is made of gas that is partially ionized by the accretion disc radiation and hence illuminates the jet with UV line emission (most prominently Lyα; Tavecchio & Ghisellini 2008). Thus, the characteristic BLR photon energy is

$$\epsilon_{\rm BLR} \sim 10\,{\rm eV}\,. \tag{90}$$

The BLR extends out to a radius

$$r_{\text{BLR}} \sim 0.1 L_{\text{d},46}^{1/2} \, \text{pc} \qquad (91)$$

where $L_{\text{d},46}$ is the luminosity of the accretion disc, L_{d}, in units of $10^{46}\,\text{erg s}^{-1}$ (Tavecchio & Ghisellini 2008; Sikora et al. 2009; Nalewajko et al. 2012). At jet propagation distances $r < r_{\text{BLR}}$ from the central engine, the jet traverses roughly isotropic ambient radiation, sourced by the BLR, of galaxy-frame energy density

$$U_{\text{BLR}} \sim \frac{L_{\text{BLR}}}{4\pi r_{\text{BLR}}^2 c} \sim 6 \times 10^{-3} \, \text{erg cm}^{-3} \qquad (92)$$

(Tavecchio & Ghisellini 2008; Sikora et al. 2009; Nalewajko et al. 2012). Here we assume that the BLR intercepts and reprocesses a certain fraction (2 per cent; Tavecchio et al. 2011) of the accretion disc light, and, therefore, that $L_{\text{BLR}} \propto L_{\text{d}}$. Thus, U_{BLR} is insensitive to L_{d}.

Farther removed from the nucleus than the BLR is the HDR, which, radiatively heated by the accretion disc, shines a quasi-thermal spectrum of temperature $T_{\text{HDR}} \simeq 1200\,\text{K}$ on to the jet. Thus (see Nenkova et al. 2008a, b; Nalewajko et al. 2012),

$$\epsilon_{\text{HDR}} \sim 3k_{\text{B}} T_{\text{HDR}} = 0.3\,\text{eV} \, . \qquad (93)$$

The HDR extends out to a distance from the central engine of

$$r_{\text{HDR}} \sim 4 \, L_{\text{d},46}^{1/2} T_{\text{HDR},3}^{-2.6} \, \text{pc} \, , \qquad (94)$$

where $T_{\text{HDR},3} \equiv T_{\text{HDR}}/1000\,\text{K} = 1.2$ (Nenkova et al. 2008a, b; Sikora et al. 2009). Thus, when $r < r_{\text{HDR}}$, the HDR radiation energy density traversed by the jet is roughly isotropic and approximately

$$U_{\text{HDR}} \sim \frac{L_{\text{HDR}}}{4\pi r_{\text{HDR}}^2 c} \sim 9 \times 10^{-5} \, \text{erg cm}^{-3} \, . \qquad (95)$$

Here, we again take a fixed fraction (in this case 10 per cent; Malmrose et al. 2011) of the disc radiation to be reprocessed by the circumnuclear structure. Thus U_{HDR}, like U_{BLR}, lacks L_{d}-dependence.

We posit that at least some quiescent and flaring blazar gamma-ray observations are powered by magnetic reconnection (Giannios, Uzdensky & Begelman 2009; Nalewajko et al. 2011; Giannios 2013; Sironi et al. 2015; Petropoulou, Giannios & Sironi 2016; Nalewajko et al. 2018; Werner et al. 2018; Christie et al. 2019; Ortuño-Macías & Nalewajko 2020; Sobacchi, Nättilä & Sironi 2021), either induced by macroscopic field polarity reversals (e.g. Giannios & Uzdensky 2019; Sironi, Rowan & Narayan 2021) or taking place at the small-scale terminus of a turbulent cascade (e.g. Zhdankin et al. 2013, 2020; Comisso & Sironi 2018, 2019; Boldyrev & Loureiro 2020; Loureiro & Boldyrev 2020; Nättilä & Beloborodov 2020; Sobacchi et al. 2021).

The blazar emission zone – the distance r where most of the emission is produced – is an important but difficult-to-constrain quantity in blazar research: by nature of being collimated along the observing line of sight, blazar jets appear point-like on the sky, and so it is not possible to deduce r directly from observations. We therefore allow r to vary over an appreciable range. If $r < r_{\text{BLR}}$, the BLR dominates the ambient radiation bathing the jet and the emission-powering reconnection zone; if $r_{\text{BLR}} < r < r_{\text{HDR}}$, the HDR dominates.[5] We consider only blazar zones far enough from the central engine ($r \gtrsim 0.01\,\text{pc}$; Dermer & Schlickeiser 2002; Sikora et al.

2009; Nalewajko et al. 2014) that the direct accretion disc light, which illuminates the jet from behind, is redshifted in the jet rest frame to a lower energy density than the BLR and HDR radiation fields. (Unlike the accretion disc light, the BLR and HDR photons impinge quasi-isotropically on the jet in the galaxy frame and are thus blueshifted when boosted to the jet frame). We do not consider $r \gg r_{\text{HDR}}$.

The BLR and HDR choke gamma-rays above the energies

$$\epsilon_{\text{c,BLR}} = \frac{(m_e c^2)^2}{\epsilon_{\text{BLR}}} \sim 30\,\text{GeV} \qquad (96)$$

and

$$\epsilon_{\text{c,HDR}} = \frac{(m_e c^2)^2}{\epsilon_{\text{HDR}}} \sim 0.9\,\text{TeV} \, , \qquad (97)$$

respectively, since the pair-production optical depths suffered by a photon traversing these regions are

$$\tau_{\text{BLR}} = \frac{U_{\text{BLR}}\sigma_{\text{T}} r_{\text{BLR}}}{5\epsilon_{\text{BLR}}} \sim 15 \qquad (98)$$

and

$$\tau_{\text{HDR}} = \frac{U_{\text{HDR}}\sigma_{\text{T}} r_{\text{HDR}}}{5\epsilon_{\text{HDR}}} \sim 300 \, . \qquad (99)$$

Note that these optical depths are *not* equal to those, $\tau_{\gamma\gamma,\text{BLR}}$ and $\tau_{\gamma\gamma,\text{HDR}}$, of the corresponding reconnection sites where the emission is sourced. Those optical depths are evaluated later using the smaller size of the emission region and with the seed photon number density U_i/ϵ_i ($i = \text{BLR}$ or HDR) boosted to the jet frame. Since quiescent FSRQs are generally not observed at very high energies ($\gtrsim 0.1\,\text{TeV}$), the most relevant model for them may be irradiation by the BLR. However, FSRQ flares are sometimes observed at up to several hundred GeV (MAGIC Collaboration 2008; Aleksić et al. 2011; H. E. S. S. Collaboration 2013; Abeysekara et al. 2015; Ahnen et al. 2015; Sitarek et al. 2015) – still below $\epsilon_{\text{c, HDR}}$. In these cases, emission between r_{BLR} and r_{HDR}, where the external photons come from the HDR, is most likely.

5.2.2 FSRQ jet parameters

In order to estimate reconnection parameters, we must now make a number of further assumptions about the nature of the jet: including its speed, shape, and magnetic field strength.

We assume that reconnection occurs in the rest frame of the jet and that the jet travels with relativistic bulk Lorentz factor $\Gamma_{\text{j}} \gg 1$ with respect to the host galaxy. We denote jet-frame quantities with primes. However, we exclude individual particle Lorentz factors (including cold magnetizations $\sigma_{\text{c},0}$) from this convention, writing them exclusively in the jet frame and without primes. We adopt a relatively high jet Lorentz factor $\Gamma_{\text{j}} = 40$. This is the same as in our earlier model (Mehlhaff et al. 2020) of the rapidly variable very high-energy ($\gtrsim 0.1\,\text{TeV}$) flare from PKS 1222+21 (Aleksić et al. 2011). A large Γ_{j} may be reconciled with more typical Lorentz factors (e.g. $\Gamma_{\text{j}} \simeq 10$) by invoking a structured jet (e.g. Ghisellini, Tavecchio & Chiaberge 2005; Begelman et al. 2008; Sikora, Rutkowski & Begelman 2016; Tavecchio & Ghisellini 2016; Tavecchio 2017; Sironi et al. 2021) in which an inner spine region, moving quickly ($\Gamma_{\text{j}} \simeq 40$) and carrying the reconnection current sheet, is surrounded by a slower moving sheath. This explicit structure is not necessary to our ensuing discussion, however.[6]

---

[5]In reality, the BLR intensity falls off smoothly with r and thus dominates to distances slightly exceeding r_{BLR} (but still much less than r_{HDR}; e.g. Nalewajko et al. 2014).

[6]Unlike in our earlier model (Mehlhaff et al. 2020), we do not entertain the sheath as a potential source of Compton seed photons here.
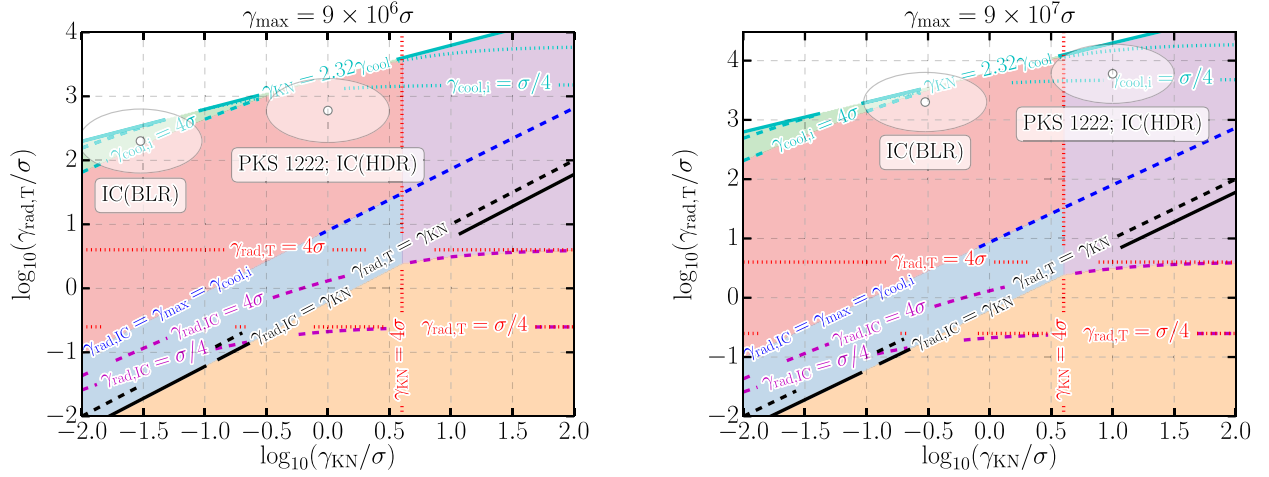
Figure 17. Radiative reconnection phase diagrams for the case $\gamma_{\rm max} = 9 \times 10^{10}$ (equation 101) and two different values of $\sigma_{\rm c,0}$ (abbreviated as σ in the plot): $\sigma_{\rm c,0} = 10^4$ (left) and $\sigma_{\rm c,0} = 10^3$ (right). Region colourings have the same meaning as in Fig. 8. In each diagram, white points indicate the parameters of equations (102)–(105) and (109)–(112), where gamma-ray emission comes primarily through Comptonization of BLR photons [IC(BLR)] and HDR photons [IC(HDR)], respectively. Larger shaded white ellipsoids indicate one order-of-magnitude uncertainties in each direction. The IC(HDR) parameters are particularly relevant to the rapid very high-energy flare from PKS 1222+21 (Aleksić et al. 2011) discussed in our earlier work (Mehlhaff et al. 2020). Because the three parameters $\gamma_{\rm rad,T}$, $\gamma_{\rm KN}$, and $\gamma_{\rm max}$ are all proportional to $\Gamma_{\rm j}^{-1}$, changing $\Gamma_{\rm j}$ is equivalent to changing $\sigma_{\rm c,0}$ by the same factor.

We imagine that the jet is conical with opening angle $\theta_{\rm j} \sim 1/5\Gamma_{\rm j}$ (Pushkarev et al. 2009) and model the comoving magnetic field strength by assuming

$$B_0'(r) \sim 0.1 \left(\frac{r}{1\,{\rm pc}}\right)^{-1} {\rm G} \qquad (100)$$

(Nalewajko et al. 2012; Mehlhaff et al. 2020), corresponding to a total jet Poynting flux luminosity of $(\theta_{\rm j} r)^2 (\Gamma_{\rm j}^2 B_0'^2) c \sim 1 \times 10^{43}\,{\rm erg\,s^{-1}}$ that is conserved in r and independent of $\Gamma_{\rm j}$. If the Poynting flux is instead dissipated (as indeed a reconnection scenario suggests), then the scaling of B_0' with r should be steeper, but perhaps not by much. For example, the fiducial striped jet model of Giannios & Uzdensky (2019), which includes magnetic dissipation through reconnection, gives a power law close to $B_0' \propto r^{-5/4}$.

We note that the typical isotropic FSRQ luminosities $L_{\rm iso} \sim 10^{48}\,{\rm erg\,s^{-1}}$ require minimum intrinsic jet power, presumably carried by magnetic fields in a reconnection scenario, of $L_{\rm iso}/\Gamma_{\rm j}^2 \sim 6 \times 10^{44}\,{\rm erg\,s^{-1}}$. We have checked that increasing our fiducial magnetic field strength [e.g. $B_0'(1\,{\rm pc}) \sim 1\,{\rm G}$] to supply such jet power does not affect our conclusions in this section, but we adopt (100) to maintain continuity with our previous work (Mehlhaff et al. 2020).

5.2.3 Reconnection parameters in FSRQs

We now explicitly estimate parameters for reconnection in an FSRQ jet illuminated by either the BLR or the dusty torus.

First, the maximum (system-size-limited) Lorentz factor is

$$\gamma_{\rm max} \equiv \frac{0.1L}{\rho_0} \sim \frac{0.1 e B_0'(r) r \theta_{\rm j}}{m_e c^2}$$
$$\sim 9 \times 10^{10} \left(\frac{\Gamma_{\rm j}}{40}\right)^{-1}. \qquad (101)$$

Note that this is independent of r because we take the reconnection layer length L to be comparable to the transverse width $r\theta_{\rm j}$ of the jet, and thus the dependence of L on r cancels against that of the magnetic field strength in equation (100).

Next we discuss the jet magnetization. In lieu of estimating a fiducial jet electron number density (which is needed to estimate $\sigma_{\rm c,0}$), we take inspiration from previous studies in which the jet is moderately magnetized at the parsec scale (e.g. Giannios 2013; Giannios & Uzdensky 2019). If the jet is mass-dominated by protons, this still allows individual electrons/positrons to attain high Lorentz factors. Assuming an electron–proton jet – and that our main results from earlier sections carry over to electron–proton plasmas containing ultrarelativistic initial magnetic energy per electron – we adopt a cold *ion* magnetization (in the jet frame) of 5, which corresponds to a fiducial jet-frame cold *electron* magnetization $\sigma_{\rm c,0} \sim 10^4$. To illustrate our high degree of uncertainty in this quantity, in Fig. 17, we also present the case $\sigma_{\rm c,0} \sim 10^3$, which corresponds to a cold ion magnetization of 0.5 (or of 5 but with 10 positrons per proton in the presence of mixed composition). In addition, we crudely assume that the cold magnetization does not change in r, which is consistent with Poynting flux conservation if one also has particle flux conservation [$\pi (r\theta_{\rm j})^2 n_0(r) = {\rm constant}$].

Now we estimate the key radiative parameters. For reconnection illuminated by the BLR ($r < r_{\rm BLR}$), we have, plugging the jet-frame seed photon energy $\epsilon_{\rm BLR}' \sim \Gamma_{\rm j}\epsilon_{\rm BLR}$ and energy density $U_{\rm BLR}' \sim \Gamma_{\rm j}^2 U_{\rm BLR}$ (with $\epsilon_{\rm BLR}$ and $U_{\rm BLR}$ given in equations 90 and 92) into definitions (8), (9), (12), and (36):

$$\gamma_{\rm rad,T}^{\rm (BLR)} \sim 2 \times 10^6 \left(\frac{r}{0.1\,{\rm pc}}\right)^{-1/2} \left(\frac{\Gamma_{\rm j}}{40}\right)^{-1}, \qquad (102)$$

$$\gamma_{\rm cool,BLR} \sim 60 \left(\frac{r}{0.1\,{\rm pc}}\right)^{-1} \left(\frac{\Gamma_{\rm j}}{40}\right)^{-1}, \qquad (103)$$

$$\gamma_{\rm KN,BLR} \sim 300 \left(\frac{\Gamma_{\rm j}}{40}\right)^{-1}, \quad {\rm and} \qquad (104)$$

$$\tau_{\gamma\gamma,\rm BLR} \sim 3 \left(\frac{r}{0.1\,{\rm pc}}\right). \qquad (105)$$

The observed photon energies emitted by particles with these radiative Lorentz factors are

$$\epsilon_{\rm obs}^{\rm (BLR)}\left(\gamma_{\rm rad,T}^{\rm (BLR)}\right) \sim \frac{1}{2}\Gamma_{\rm j}\gamma_{\rm rad,T}^{\rm (BLR)} m_e c^2 \sim 20 \left(\frac{r}{0.1\,{\rm pc}}\right)^{-1} {\rm TeV}, \qquad (106)$$

$$\epsilon_{\text{obs}}^{(\text{BLR})}\left(\gamma_{\text{cool,BLR}}\right) \sim \frac{4}{3}\Gamma_{\text{j}}^2\gamma_{\text{cool,BLR}}^2\epsilon_{\text{BLR}}$$

$$\sim 80\left(\frac{r}{0.1\,\text{pc}}\right)^{-2}\text{MeV},\quad\text{and} \qquad (107)$$

$$\epsilon_{\text{obs}}^{(\text{BLR})}\left(\gamma_{\text{KN,BLR}}\right) \sim \frac{4}{3}\Gamma_{\text{j}}^2\gamma_{\text{KN,BLR}}^2\epsilon_{\text{BLR}} \sim 2\,\text{GeV}. \qquad (108)$$

These photon energies are independent of the jet Lorentz factor Γ_{j}. One of them, $\epsilon_{\text{obs}}^{(\text{BLR})}(\gamma_{\text{KN,BLR}}) \sim 2\,\text{GeV}$, depends on no unknowns and implies that virtually all the $> 1\,\text{GeV}$ emission, routinely observed by *Fermi* (Ghisellini 2011), is emitted in the Klein–Nishina regime (if Comptonized from BLR photons). Note that, in evaluating the photon energies emitted by $\gamma = \gamma_{\text{KN}}$ particles, one may use either the Thomson scaling $4\Gamma_{\text{j}}^2\gamma_{\text{KN}}^2\epsilon_{\text{ph}}/3$ (as we have done) or the Klein–Nishina scaling $\Gamma_{\text{j}}\gamma_{\text{KN}}m_e c^2/2$; this choice modifies the estimate by less than a factor of 2.

If, instead, $r_{\text{BLR}} < r < r_{\text{HDR}}$, the radiation from the dusty torus is strongest, yielding, from the HDR estimates (93) and (95),

$$\gamma_{\text{rad,T}}^{(\text{HDR})} \sim 6\times10^6\left(\frac{r}{1\,\text{pc}}\right)^{-1/2}\left(\frac{\Gamma_{\text{j}}}{40}\right)^{-1}, \qquad (109)$$

$$\gamma_{\text{cool,HDR}} \sim 400\left(\frac{r}{1\,\text{pc}}\right)^{-1}\left(\frac{\Gamma_{\text{j}}}{40}\right)^{-1}, \qquad (110)$$

$$\gamma_{\text{KN,HDR}} \sim 1\times10^4\left(\frac{\Gamma_{\text{j}}}{40}\right)^{-1},\quad\text{and} \qquad (111)$$

$$\tau_{\gamma\gamma,\text{HDR}} \sim 20\left(\frac{r}{1\,\text{pc}}\right). \qquad (112)$$

The corresponding observed photon energies are

$$\epsilon_{\text{obs}}^{(\text{HDR})}\left(\gamma_{\text{rad,T}}^{(\text{HDR})}\right) \sim \frac{1}{2}\Gamma_{\text{j}}\gamma_{\text{rad,T}}^{(\text{HDR})}m_e c^2 \sim 60\left(\frac{r}{1\,\text{pc}}\right)^{-1}\text{TeV}, \qquad (113)$$

$$\epsilon_{\text{obs}}^{(\text{HDR})}\left(\gamma_{\text{cool,HDR}}\right) \sim \frac{4}{3}\Gamma_{\text{j}}^2\gamma_{\text{cool,HDR}}^2\epsilon_{\text{HDR}}$$

$$\sim 100\left(\frac{r}{1\,\text{pc}}\right)^{-2}\text{MeV},\quad\text{and} \qquad (114)$$

$$\epsilon_{\text{obs}}^{(\text{HDR})}\left(\gamma_{\text{KN,HDR}}\right) \sim \frac{4}{3}\Gamma_{\text{j}}^2\gamma_{\text{KN,HDR}}^2\epsilon_{\text{HDR}} \sim 70\,\text{GeV}. \qquad (115)$$

The photon energy $\epsilon_{\text{obs}}^{(\text{HDR})}(\gamma_{\text{KN,HDR}}) \sim 70\,\text{GeV}$ [like the corresponding energy $\epsilon_{\text{obs}}^{(\text{BLR})}(\gamma_{\text{KN,BLR}})$ for Comptonization within the BLR] does not depend on any unknowns and implies that IC(HDR) TeV emission from FSRQs is produced almost entirely in the Klein–Nishina regime. Irrespective of the dominant radiation field (BLR or HDR), the overall IC radiative cut-off energy, $\gamma_{\text{rad,IC}}$, is effectively infinite, scaling according to equation (30) as $\gamma_{\text{rad,IC}} \propto \gamma_{\text{KN}}\exp[(2/9)(\gamma_{\text{rad,T}}/\gamma_{\text{KN}})^2]$.

Fig. 17 displays the BLR (102–105) and HDR (109–112) energy scales by plotting them as points in a radiative reconnection 'phase diagram' in the style of Fig. 8. This clearly illustrates the main result of this section: *reconnection in FSRQ jets proceeds in the regime governed by Klein–Nishina and pair-production physics*. This does not depend on whether the dominant seed photons stem from the BLR or the HDR. Broadly speaking (we discuss caveats and technical points below), reconnection proceeds in the red region. Here, X-point acceleration is not inhibited by radiative losses ($\gamma_{\text{rad,IC}} \gg \gamma_{\text{max}} \gg \gamma_{\text{X}}$), but subsequently most, if not all, particles cool strongly – on time-scales shorter than L/c – rendering reconnection efficiently radiative. Furthermore, because $\gamma_{\text{X}} > \gamma_{\text{KN}}$, many particles become impulsively energized up to energies where they radiate in the Klein–Nishina limit, producing above-threshold radiation that may activate pair feedback (Section 4).

Before moving to additional observational implications, we discuss a few secondary technical details to flesh out this basic picture. First, though one might worry that the IC(HDR) model, depending on $\sigma_{\text{c,0}}$, crosses into the purple region of Fig. 17 where the nominal intrinsic X-point acceleration cut-off energy, $\gamma_{\text{X}} \simeq 4\sigma_{\text{c,0}}$, falls below γ_{KN}, this is not a huge concern. The observed photon energies $\epsilon_{\text{obs}}^{(\text{BLR})}(\gamma_{\text{KN,BLR}}) \sim 2\,\text{GeV}$ and $\epsilon_{\text{obs}}^{(\text{HDR})}(\gamma_{\text{KN,HDR}}) \sim 70\,\text{GeV}$ are routinely observed in quiescent FSRQ spectra (for the former; Ghisellini 2011; Madejski & Sikora 2016) and in TeV outbursts (for the latter; MAGIC Collaboration 2008; Aleksić et al. 2011; H. E. S. S. Collaboration 2013; Abeysekara et al. 2015; Ahnen et al. 2015; Sitarek et al. 2015). Thus, in a reconnection scenario, observations, in fact, suggest that γ_{X}, whether through high-$\sigma_{\text{c,0}}$ or through circumventing the conventional $4\sigma_{\text{c,0}}$-limit (as discussed in Section 5.1.2), is high enough to accelerate particles beyond γ_{KN}, pushing reconnection into the Klein–Nishina radiative regime.

Next, one should note that the pair-production optical depth of the BLR-illuminated reconnection layer is not very large: $\tau_{\gamma\gamma,\text{BLR}} \sim 3$. According to equation (33), this means that even the most strongly radiative particles possess cooling times $t_{\text{cool,IC}}(\gamma) \sim L/c\tau_{\gamma\gamma}$ only marginally faster than a dynamical time. This somewhat limits the overall radiative efficiency of reconnection, as signalled by one of the IC(BLR) points in Fig. 17 bordering on the green region where Klein–Nishina suppression of the IC cross-section begins to shut down radiative cooling for the highest energy ($\gamma > \gamma_{\text{cool,2}}$) particles. Plus, since $\tau_{\gamma\gamma} < 10$, some newborn pairs may escape the system before being swept into the layer (f_{noesc} drops below unity; see Section 4.1), which diminishes the likelihood of swing cycles for the IC(BLR) scenario.

Regarding the multiplicity of produced pairs (Section 4.3), the estimates of this section yield $\gamma_{\text{KN, BLR}}/\sigma_{\text{c,0}} \geq 0.03$ and $\gamma_{\text{KN, HDR}}/\sigma_{\text{c,0}} \geq 1$ for $\sigma_{\text{c,0}} \leq 10^4$ and $\Gamma_{\text{j}} \leq 40$. Thus, the requirement $\gamma_{\text{KN}} > \sigma_{\text{c,0}}/10$ to ensure a pair multiplicity (the ratio of number densities of newborn pairs to originally present pairs), η, less than unity is roughly satisfied. The BLR may marginally support $\eta \gtrsim 1$ because $\max(\eta) \simeq \sigma_{\text{c,0}}/10\gamma_{\text{KN}}$ (Section 4.3). However this is doubtful for several reasons: (i) $\gamma_{\text{KN, BLR}}/\sigma_{\text{c,0}} = 0.03$ is a lower bound (increasing for smaller $\sigma_{\text{c,0}}$ or Γ_{j}); (ii) η may never attain its global maximum; and (iii) the result $\max(\eta) \simeq \sigma_{\text{c,0}}/10\gamma_{\text{KN}}$ ignores particle escape, which may be an issue for reconnection inside the BLR, as mentioned in the previous paragraph. Thus, our estimates support the picture of tenuous newborn pairs that control the energy density of the inflow plasma but not the inflowing particle count.

5.2.4 Relevance to FSRQ observations

Next, we comment on a few observational issues that our model of pair-regulated Klein–Nishina reconnection may help to address. First, and as discussed in Section 5.1.2, Klein–Nishina reconnection may enhance the range of photon energies that are kinetically beamed. This provides an attractive explanation for rapid very high-energy FSRQ flares (of the kind first observed by Aleksić et al. 2011). A kinetic beaming scenario has already been advocated on energy-budget grounds by Nalewajko et al. (2012), and has recently been explored by us in the context of Thomson radiative reconnection (Mehlhaff et al. 2020). However, our earlier work requires a dense population of seed photons – much denser than those supplied by the BLR and HDR – in order to facilitate the high degree of radiative efficiency needed by kinetic beaming. Therefore, Mehlhaff et al. (2020) disfavour a single-zone external IC(HDR) scenario and suggest a two-zone (e.g. spine-sheath) configuration where the seed

photons are supplied by the jet itself. However, if Klein–Nishina effects ease the radiative efficiency requirement, an IC(HDR) model may still be compatible with kinetic beaming.

Next, although we would like to refrain from trying to make detailed spectral interpretations, we note that the spectra of flaring FSRQs in the very high energy (VHE; $\gtrsim 0.1$ TeV) band are generally quite steep, with intrinsic (deabsorbed) photon number index $\Gamma_{\mathrm{VHE}} \gtrsim 2.5$ (MAGIC Collaboration 2008; Aleksić et al. 2011; Ahnen et al. 2015). If one assumes that there is no significant spectral break between the Thomson and deep Klein–Nishina regimes (Moderski et al. 2005), then $\Gamma_{\mathrm{VHE}} \simeq (p_{\mathrm{T}}+1)/2 \simeq (p+2)/2$. This suggests a steep injected index $p \gtrsim 3$, which is realized in our reconnection 'equation of state' $p(\sigma_{\mathrm{h}}) = 1 + 2/\sqrt{\sigma_{\mathrm{h}}}$, equation (50), if $\sigma_{\mathrm{h},0} \lesssim 1$. This may occur when the ions are moderately magnetized and dominate the initial hot magnetization (e.g. Werner et al. 2018). In this situation, pair feedback, because it cannot reduce σ_{h} to below order unity (equation 49), would probably not strongly modify the spectrum.

However, if $\sigma_{\mathrm{h},0} \gg 1$, pair feedback may produce the observed steep spectral indices during dramatic swing cycles (e.g. Fig. 15), wherein the observable emission is dominated by the layer (Section 5.1.1) in the low magnetization $\sigma_{\mathrm{h},<} \sim 1$ part of the cycle. The requirement $\sigma_{\mathrm{h},0} \gg 1$ could be realized if either: (i) the initial hot magnetization governing the distribution of accelerated electrons/positrons decouples from the (order unity) ion magnetization, or (ii) a highly magnetized pair-dominated jet region furnishes the upstream material for reconnection.

In addition, to support swing cycles, the layer generally requires high energy recapture efficiency $\xi \simeq f_{\mathrm{nocool}} \sim 1$ (e.g. Fig. 15). In Appendix B, we show that order-unity energy retention factors f_{nocool} require a large cut-off $z = \gamma_2/\gamma_{\mathrm{KN}}$ (e.g. $z \gtrsim 1000$ in Fig. B3). Such high cut-offs might not be realized because the observed Compton dominance in FSRQs is often $U_{\mathrm{ph}}/U_{\mathrm{B}} \sim 100$, suggesting that the particle energy, γ_{s}, beyond which synchrotron losses outcompete IC losses and may hence limit further particle acceleration, is about $\gamma_{\mathrm{s}} \sim 30\gamma_{\mathrm{KN}}$ (equation 28).[7]

However, there are a few effects that may serve to promote swing cycles in spite of non-zero synchrotron losses. First, the most energetic particles responsible for emitting pair-producing photons are likely accelerated near reconnection X-points deep inside the current layer. There, the magnetic field is weaker (Uzdensky et al. 2011; Cerutti et al. 2012a), which reduces synchrotron radiation relative to IC losses, potentially allowing the latter to remain dominant to higher energies than γ_{s}. A second possibility is that swing cycles actually set in at lower values of z than we predict. This is because we conservatively estimate a steep distribution of pair-producing photons penetrating the upstream plasma $\Gamma = p + 0.5$ in Section 4.2 (rather than, e.g. $\Gamma = p$ as in Moderski et al. 2005), and therefore may artificially underestimate f_{nocool}. Finally, and related to the previous point, we find (not presented here) that the threshold on z for swing cycles to kick in is quite sensitive to the precise dependence of f_{nocool} on z. These considerations demonstrate the need for radiative kinetic simulations to examine in detail whether swing cycles are possible, when they occur, and how they impact the spectrum of observed radiation.

As a completely separate prospect, our model may connect to the matter-antimatter balance in FSRQ jets, which is difficult to constrain from observations alone (Madejski & Sikora 2016). This ratio may be initially imprinted at the base of the jet – for example, by magnetospheric spark-gap discharges (e.g. Blandford & Znajek 1977; Beskin, Istomin & Parev 1992; Chen, Yuan & Yang 2018; Ford, Keenan & Medvedev 2018; Crinquand et al. 2020) or by interaction between the nascent jet and the accretion flow (e.g. Ripperda, Bacchini & Philippov 2020; Wong, Ryan & Gammie 2021) – but it may also be modified *in situ* as the jet propagates. Although we expect relatively few newborn pairs compared to the initial number of upstream particles in each reconnection episode (Section 4.3), continuous, repeated reconnection occurring as the jet ploughs through ambient radiation fields could still lead to secular growth in the number of pairs present in the plasma. This may be relevant to FSRQs in their quiescent states, where their spectral cut-offs are broadly consistent with gamma-ray absorption in the BLR. Furthermore, because $\tau_{\mathrm{BLR}} > 1$, even if the reconnection region itself is optically thin ($\tau_{\gamma\gamma,\mathrm{BLR}} < 1$), emitted gamma-rays may still produce pairs in the jet before exiting the BLR (though they do not strongly impact the reconnection dynamics).

In summary, reconnection in FSRQ jets illuminated by either the BLR or the dusty torus is expected to occur in the radiative regime governed by pair-production and Klein–Nishina physics. Here, kinetic beaming and pair-feedback-initiated swing cycles may explain the time-scales of rapid TeV flares and typical FSRQ VHE spectral indices, respectively. However, confirming whether these mechanisms are active in Klein–Nishina radiative reconnection depends on several unknowns and remains a key open question for future simulations to address. In addition, *in situ* pair production driven by reconnection could contribute to the pair content of quiescent FSRQ jets.

5.2.5 Reconnection parameters in black hole ADCe

Magnetic reconnection may also power emission from a highly magnetized corona sandwiching a black hole accretion disc (Liang & Price 1977; Galeev, Rosner & Vaiana 1979; Di Matteo 1998; Goodman & Uzdensky 2008; Uzdensky & Goodman 2008; Uzdensky 2016; Beloborodov 2017; Werner et al. 2019; Sironi & Beloborodov 2020). Here, we focus on the high/soft states of black hole X-ray binaries, showing that coronal reconnection in this context likely proceeds in a highly radiative Klein–Nishina regime. We term as high/soft any state where the quasi-thermal (~ 1 keV) spectral component – which is attributed to an optically thick, geometrically thin accretion disc (Shakura & Sunyaev 1973) – strongly dominates over the non-thermal component, which extends to much higher, hard X- and gamma-ray energies (e.g. Remillard & McClintock 2006). Because reconnection intrinsically gives rise to a non-thermal distribution of particle energies, a reconnection scenario does not rely on repeated scatterings by an isothermal plasma of moderate optical depth (e.g. Eardley & Lightman 1976; Shapiro, Lightman & Eardley 1976; Rybicki & Lightman 1979) to explain the observed non-thermal spectrum.

The bright quasi-thermal disc emission in the high/soft state provides an intense soft photon bath to the coronal reconnection region, and we assume that Comptonization of disc photons is responsible for the observed non-thermal spectrum. We now estimate the energy scales $\gamma_{\mathrm{KN,disc}}$ and $\gamma_{\mathrm{rad,T}}^{(\mathrm{disc})}$ associated with the disc seed photons, as well as the energy scales $\sigma_{\mathrm{c},0}$ and γ_{max}. Determining these then allows us to accomplish our main goal for this section: characterizing the radiative regime of reconnection.

---

[7]Generally, $U_{\mathrm{ph}}/U_{\mathrm{B}}$ is equated to the observed ratio of IC-to-synchrotron luminosities. This is sound in one-zone emission scenarios where Klein–Nishina effects do not suppress IC emission near the IC spectral peak. The latter requirement is generally satisfied because Compton FSRQ peaks typically fall at lower energies than $\epsilon_{\mathrm{obs}}^{(\mathrm{BLR})}(\gamma_{\mathrm{KN,BLR}})$ and $\epsilon_{\mathrm{obs}}^{(\mathrm{HDR})}(\gamma_{\mathrm{KN,HDR}})$.

Beginning with $\gamma_{\text{KN,disc}}$, the characteristic seed photon energy emitted from the disc is

$$\epsilon_{\text{disc}} \sim 1 \, \text{keV} \,, \tag{116}$$

as both theorized (Shakura & Sunyaev 1973) and observed (e.g. in Cyg X-1; McConnell et al. 2002; Remillard & McClintock 2006). Through equation (19), the Klein–Nishina Lorentz factor is then

$$\gamma_{\text{KN,disc}} \equiv \frac{m_e c^2}{4 \epsilon_{\text{disc}}} \sim 100 \,. \tag{117}$$

The corresponding energy to which photons encountering $\gamma_{\text{KN,disc}}$ particles are upscattered is

$$\epsilon_{\text{obs}}^{(\text{disc})} \left(\gamma_{\text{KN,disc}} \right) \sim 20 \, \text{MeV} \,. \tag{118}$$

Emission at these energies has been observed by the *Fermi* Large Area Telescope (LAT) from Cyg X-1 (Zdziarski et al. 2017).

We now move on to estimate $\gamma_{\text{rad,T}}^{(\text{disc})}$. This requires (equation 8) the energy density of the emission intercepted by the corona at a distance r from the black hole, which is roughly

$$U_{\text{disc}} \sim \frac{L_{\text{disc}}}{4 \pi r^2 c} \,. \tag{119}$$

Here, L_{disc} is the luminosity of the disc (emitted mostly near its inner edge) and r is assumed to exceed the innermost disc orbit.

In the high/soft state L_{disc} can be quite high – up to several per cent of the Eddington limit, $L_{\text{Edd}} = 4\pi G M_{\text{BH}} m_p c / \sigma_{\text{T}} = 4\pi r_g m_p c^3 / \sigma_{\text{T}} \simeq 1 \times 10^{39} (M_{\text{BH}}/10 \, \text{M}_\odot) \, \text{erg s}^{-1}$ (e.g. 4 per cent for Cyg X-1; McConnell et al. 2002). Furthermore, r is at least several gravitational radii, $r_g \equiv G M_{\text{BH}}/c^2 \simeq 1 \times 10^6 (M_{\text{BH}}/10 \, \text{M}_\odot) \, \text{cm}$. Thus, an upper bound and fiducial scale for the energy density of ambient radiation shining on to the reconnection layer is (cf. Beloborodov 2017)

$$U_\star \equiv \frac{L_{\text{Edd}}}{4 \pi r_g^2 c} = \frac{m_p c^2}{r_g \sigma_{\text{T}}} \simeq 2 \times 10^{15} \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right)^{-1} \text{erg cm}^{-3} \,. \tag{120}$$

Here, the fiducial black hole mass is 10 M$_\odot$. In the following, we estimate $r \sim r_g$ and $U_{\text{disc}} \sim 0.04 U_\star$, corresponding to $L_{\text{disc}} \simeq 0.04 L_{\text{Edd}}$ observed from Cyg X-1 (McConnell et al. 2002). The parameter $U_\star$ also provides a fiducial scale for the magnetic field energy density in the disc required to transport angular momentum outward (Shakura & Sunyaev 1973; Begelman, Blandford & Rees 1984; Goodman & Uzdensky 2008; Beloborodov 2017).

Beyond the fiducial scale $U_\star$, there are a number of uncertain geometric factors that control how the radiation and magnetic field energy densities decay as one moves from the vicinity of the black hole in the disc to the reconnection region in the corona. This leads to uncertainty in $\gamma_{\text{rad,T}}^{(\text{disc})}$, which depends on both U_{disc} and the coronal magnetic field energy density, $U_{\text{B,cor}}$. We parametrize these uncertainties by exhibiting, in the expressions to follow, U_{disc} as a multiple of $U_\star$ and $U_{\text{B,cor}}$ as $U_{\text{disc}}/C_{\text{d}}$, where $C_{\text{d}} \equiv U_{\text{disc}}/U_{\text{B,cor}}$ is the nominal Compton dominance. Without knowing the geometric factors governing how the seed photon and magnetic energy densities decay from the disc to the coronal reconnection region, we assume for simplicity that they fall off in the same way relative to $U_\star$, adopting a fiducial value $C_{\text{d}} = 1$. We discuss consequences on Klein–Nishina radiative reconnection if C_{d} is truly ~ 1 at the end of this section.

Following these conventions, one may write the fiducial upstream magnetic field strength as (e.g. Goodman & Uzdensky 2008; Uzden-

sky 2016; Beloborodov 2017)

$$B_0 \equiv \sqrt{8\pi U_{\text{B,cor}}} = \sqrt{8\pi U_{\text{disc}}/C_{\text{d}}}$$
$$\sim 4 \times 10^7 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right)^{1/2} \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right)^{-1/2} \left(\frac{C_{\text{d}}}{1} \right)^{-1/2} \text{G} \tag{121}$$

and then use equation (8) to write

$$\gamma_{\text{rad,T}}^{(\text{disc})} = \sqrt{\frac{0.3 e B_{0,\text{cor}}}{4 \sigma_{\text{T}} U_{\text{disc}}}} = \left(\frac{0.3 e}{4 \sigma_{\text{T}}} \sqrt{\frac{8\pi}{C_{\text{d}} U_{\text{disc}}}} \right)^{1/2}$$
$$\sim 6 \times 10^3 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right)^{-1/4} \left(\frac{C_{\text{d}}}{1} \right)^{-1/4} \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right)^{1/4} \,. \tag{122}$$

Though the complicated geometry precludes determining U_{disc} and $U_{\text{B,cor}}$ precisely, the actual value of $\gamma_{\text{rad,T}}^{(\text{disc})}$ depends only weakly on these quantities.

From equations (117) and (122), one sees that $\gamma_{\text{rad,T}}^{(\text{disc})}$ is much higher than $\gamma_{\text{KN,disc}}$. This opens up the possibility that radiative reconnection operates in the Klein–Nishina regime, since particles are not radiatively inhibited from becoming accelerated beyond $\gamma_{\text{KN,disc}}$. Additionally, the radiative particle acceleration limit is not equal to $\gamma_{\text{rad,T}}^{(\text{disc})}$, but is strongly modified by Klein–Nishina effects to $\gamma_{\text{rad,IC}}^{(\text{disc})} \gg \gamma_{\text{rad,T}}^{(\text{disc})}$ (equation 30). For reference, particles at $\gamma_{\text{rad,T}}^{(\text{disc})}$ Comptonize disc photons to energy

$$\epsilon_{\text{obs}}^{(\text{disc})} \left(\gamma_{\text{rad,T}}^{(\text{disc})} \right) \sim \frac{1}{2} \gamma_{\text{rad,T}}^{(\text{disc})} m_e c^2$$
$$\sim 2 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right)^{-1/4} \left(\frac{C_{\text{d}}}{1} \right)^{-1/4} \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right)^{1/4} \text{GeV} \,. \tag{123}$$

Next, to fully specify the radiative reconnection regime, we estimate the remaining parameters γ_{max} and $\sigma_{\text{c,0}}$. We assume that the characteristic current sheet length is

$$r_g \simeq 1 \times 10^6 \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right) \text{cm} \,. \tag{124}$$

This gives an extremely large system-size-limited Lorentz factor

$$\gamma_{\text{max}} \equiv \frac{0.1 r_g e B_0}{m_e c^2} = \frac{0.1 r_g e}{m_e c^2} \sqrt{\frac{8\pi U_{\text{disc}}}{C_{\text{d}}}}$$
$$\simeq 3 \times 10^9 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right)^{1/2} \left(\frac{C_{\text{d}}}{1} \right)^{-1/2} \left(\frac{M_{\text{BH}}}{10 \, \text{M}_\odot} \right)^{1/2} \,. \tag{125}$$

With $\gamma_{\text{KN,disc}}$, $\gamma_{\text{rad,T}}^{(\text{disc})}$, and γ_{max} estimated, $\gamma_{\text{cool,disc}}$ (equation 9) and $\tau_{\gamma\gamma,\text{disc}}$ (equation 36) follow. Respectively, they are

$$\gamma_{\text{cool,disc}} \equiv \frac{3 m_e c^2}{4 U_{\text{disc}} \sigma_{\text{T}} r_g} \sim \frac{3}{4} \frac{U_\star}{U_{\text{disc}}} \frac{m_e}{m_p}$$
$$\simeq 0.01 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right)^{-1} \tag{126}$$

and

$$\tau_{\gamma\gamma,\text{disc}} \equiv \frac{3 \gamma_{\text{KN,disc}}}{5 \gamma_{\text{cool,disc}}} \sim 8 \times 10^3 \left(\frac{U_{\text{disc}}}{0.04 U_\star (M_{\text{BH}})} \right) \,. \tag{127}$$

Neither $\gamma_{\text{cool,disc}}$ nor $\tau_{\gamma\gamma,\text{disc}}$ depend on any unknowns besides the ratio $U_{\text{disc}}/U_\star$. Unlike in FSRQ jets, the formal γ_{cool} is now small. Thus, it is not a physical Lorentz factor, but just signals that all particles cool to non-relativistic energies faster than L/c. In other words, the system is highly compact: $\ell_{\text{cor}} \sim 1/\gamma_{\text{cool,cor}} \gg 1$. For particles that have cooled to non-relativistic energies, IC scattering reduces to Thomson scattering and the observed photon energy is just $\epsilon_{\text{obs}}^{(\text{cor})}(1) = \epsilon_{\text{cor}}$.
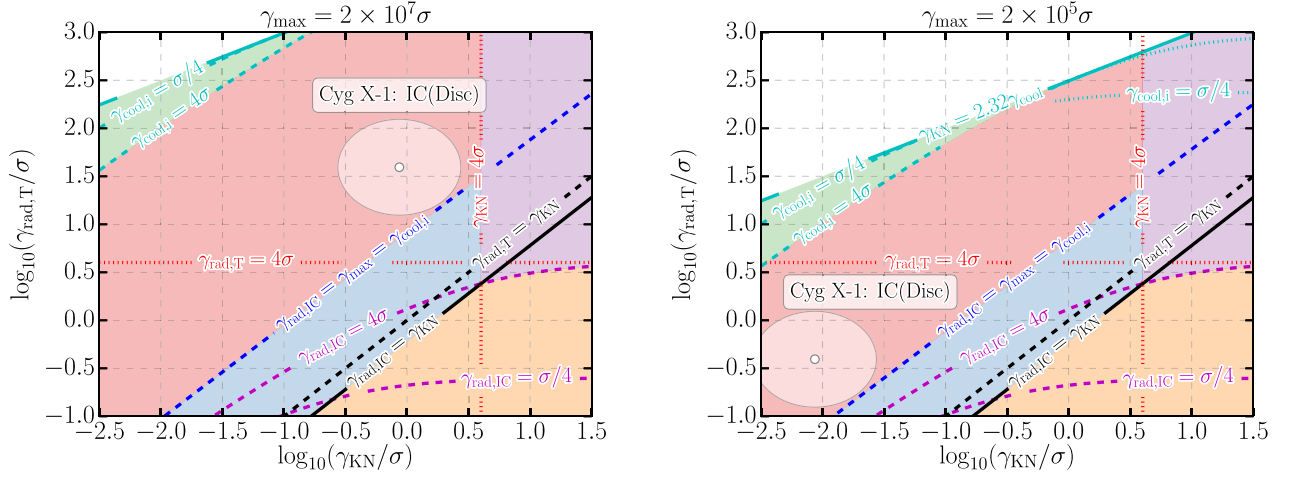
Figure 18. Radiative reconnection phase diagrams for parameters relevant to accreting black holes in their high/soft states (particularly Cyg X-1). Colour-coding is the same as in Figs 8 and 17. As in those figures, $\sigma_{c,0}$ is abbreviated as σ. White points indicate the parameters $\gamma_{\rm rad,T}^{\rm (disc)}$ and $\gamma_{\rm KN, disc}$ estimated in this section; they are surrounded by ellipses that show an arbitrarily chosen uncertainty of one decade in each direction. We display two diagrams corresponding to a range (131) of the cold magnetization $\sigma_{c,0}$ (based on typical Thomson optical depths $\tau_{\rm T}$). These are $\sigma_{c,0} = 10^2$ ($\tau_{\rm T} = 1$; left) and $\sigma_{c,0} = 10^4$ ($\tau_{\rm T} = 10^{-2}$; right). Throughout the $\sigma_{c,0}$ range, reconnection operates in a domain strongly impacted by Klein–Nishina effects and pair production.

We move now to our final estimate: the cold electron magnetization. As in the case of FSRQ jets, we are fairly uncertain of this parameter because we do not know the background coronal electron density n_0. To proceed, let us recast n_0 in terms of the Thomson optical depth $\tau_{\rm T} = n_0 \sigma_{\rm T} r_g$ along the reconnection layer (equation 124). This coincides with the optical depth of the corona itself assuming the coronal scale height is $\sim r_g$. The cold *electron* magnetization can then be compactly written in terms of $\tau_{\rm T}$ and the *magnetic compactness*,

$$\ell_{\rm B} \equiv \frac{U_{\rm B,cor}\sigma_{\rm T} r_g}{m_e c^2} = \frac{U_{\rm disc}\sigma_{\rm T} r_g}{C_{\rm d} m_e c^2} = \frac{U_{\rm disc} m_{\rm p}}{C_{\rm d} U_\star m_e}$$
$$\sim 70 \left(\frac{U_{\rm disc}}{U_\star(M_{\rm BH})}\right)\left(\frac{C_{\rm d}}{1}\right)^{-1}, \tag{128}$$

as

$$\sigma_{c,0} \equiv \frac{2U_{\rm B,cor}}{n_0 m_e c^2} \equiv 2\frac{\ell_{\rm B}}{\tau_{\rm T}}. \tag{129}$$

Thus, our uncertainty of n_0 is shifted on to $\tau_{\rm T}$. We consider a range of typical values $10^{-2} \le \tau_{\rm T} \le 1$ commonly inferred for coronae of accreting black holes, including in the high/soft state (e.g. Eardley & Lightman 1976; Poutanen, Krolik & Ryde 1997; Gierliński et al. 1999; McConnell et al. 2002; Goodman & Uzdensky 2008; Beloborodov 2017). Then, rewriting the right-hand side of (129) as

$$\sigma_{c,0} \equiv 2\frac{\ell_{\rm B}}{\tau_{\rm T}} = \frac{2U_{\rm disc} m_{\rm p}}{C_{\rm d} U_\star \tau_{\rm T} m_e} \tag{130}$$

and plugging in $10^{-2} \le \tau_{\rm T} \le 1$ gives

$$1 \times 10^2 \le \sigma_{c,0}\left(\frac{C_{\rm d}}{1}\right)\left(\frac{U_{\rm disc}}{0.04 U_\star(M_{\rm BH})}\right)^{-1} \lesssim 1 \times 10^4. \tag{131}$$

The corresponding range in the number density is $n_0 = \tau_{\rm T}/\sigma_{\rm T} r_g \in [10^{16}\,{\rm cm}^{-3}, 10^{18}\,{\rm cm}^{-3}]$ (cf. Goodman & Uzdensky 2008).

We display the estimates in this section in a radiative reconnection phase diagram, as done for FSRQs earlier, in Fig. 18. This illustrates the main point of all these estimates: *coronal reconnection in the high/soft states of accreting black holes is likely strongly impacted by Klein–Nishina and pair-production physics*. For $\tau_{\rm T} \in [10^{-2}, 1]$, reconnection occurs in the red region of the phase diagram, where $\gamma_{\rm rad,IC} > \gamma_{\rm max} > \gamma_{\rm cool,2} > 4\sigma_{c,0} > \gamma_{\rm KN} > \gamma_{\rm cool,1}$. Thus, ra-

diative losses do not hinder acceleration near reconnection X-points, but virtually all particles are still cooled strongly (on time-scales much shorter than L/c). In addition, a broad distribution of particles develops at energies $\gg \gamma_{\rm KN}$. These particles emit above-threshold photons that then produce pairs which may feedback on reconnection.

We now mention a few subtler details associated with this picture. First, we note that our model of pair-regulated Klein–Nishina reconnection in black hole ADCe is complimentary to the reconnection scenario laid out in the context of the low/hard states of black hole X-ray binaries by Beloborodov (2017). Pair production also features in that earlier work, but, there, it occurs when two layer-Comptonized photons, both carrying energy $< m_e^2 c^4/\epsilon_{\rm ph}$, collide. In contrast, in this study, a pair is produced when a single Comptonized photon with even higher energy, $> m_e^2 c^4/\epsilon_{\rm ph}$, is absorbed by the seed radiation field.

Next, we mention the implications of our estimates for the pair multiplicity, η. Equations (117) and (131) show that the requirement $\gamma_{\rm KN,cor}/\sigma_{c,0} \ge 0.1$ for η to be less than unity (Section 4.3) is not always satisfied. Thus, the pairs sourced from gamma-ray absorption by the disc seed photons may sometimes reach a number density comparable to the background upstream plasma. However, a stronger nominal Compton dominance $C_{\rm d} \equiv U_{\rm disc}/U_{\rm B,cor}$ drives down $\sigma_{c,0}$ (equation 131) and pushes reconnection into the regime where the produced pairs are tenuous.

Lastly, we discuss the implications of the assumption $C_{\rm d} \sim 1$. As mentioned earlier, $U_{\rm disc}$ and $U_{\rm B,cor}$ are both set, modulo complicated geometric factors, by the scale $U_\star$ (or, in our specific estimates, by $0.04 U_\star$). Thus, it is possible that $C_{\rm d}$ may be close to unity. However, our model of pair-regulated Klein–Nishina reconnection (Section 4) ignores synchrotron losses, and so, in the presence of such modest $C_{\rm d}$, synchrotron cooling may need to be suppressed in order for the picture of Compton cooling and pair regulation to remain valid. There are several ways for this to occur. Near reconnection X-points, the magnetic field is weaker, and so synchrotron cooling is reduced there (Uzdensky 2011; Cerutti et al. 2012a; see discussion of FSRQs). In addition, as discussed in Appendix C, the angular distribution of particles radiating above-threshold photons from the layer may cause the birth velocities of newborn upstream pairs to be nearly parallel to the reconnection mid-plane. Because the

upstream magnetic field is also parallel to this plane, the pitch angles of such newborn pairs may be small, allowing their cooling to be IC dominated. Finally, synchrotron self-absorption may inhibit synchrotron cooling among the lower energy particles (e.g. with $\gamma \lesssim 10$; Beloborodov 2017).

However, if $C_d > 1$, then our estimates numerically change, but the ordering of scales (region of the radiative phase diagram occupied by reconnection) remains essentially the same. For reference, $\gamma_{KN}/\sigma_{c,0} \propto C_d$; $\gamma_{rad,T}/\sigma_{c,0} \propto C_d^{3/4}$; and $\gamma_{max}/\sigma_{c,0} \propto C_d^{1/2}$. If, for illustration, one scales up C_d to 10, the ordering $\gamma_{rad,IC} > \gamma_{max} > \gamma_{cool,2} > \gamma_{rad,T} > 4\sigma_{c,0} > \gamma_{KN}$ is preserved except at the high-$\sigma_{c,0}$ (low-τ_T) end of the interval (131). There, γ_{KN} switches places with $4\sigma_{c,0}$, which brings reconnection into the magenta region (e.g. Fig. 18) of the phase diagrams where the nominal impulsive X-point acceleration limit, $4\sigma_{c,0}$, falls below the minimum energy, γ_{KN}, for Klein–Nishina effects kick in. However, like for FSRQs (Section 5.2.3), we may infer from observations (equation 118) that particles are accelerated at least up to γ_{KN}, suggesting that, if $4\sigma_{c,0} < \gamma_{KN}$, the nominal $4\sigma_{c,0}$ X-point acceleration limit is circumvented and that reconnection may still be strongly influenced by Klein–Nishina radiative physics.

5.2.6 Relevance to observations of black hole X-ray binaries

We now describe the implications of our reconnection model for high/soft states of black hole ADCe in accretion-powered X-ray binaries. We are aware of two such systems for which gamma-ray observations have been reported: Cyg X-1 (detected by the *Fermi* LAT, Zdziarski et al. 2017) and Cyg X-3 (detected by *AGILE*, Tavani et al. 2009, and by the *Fermi* LAT, Fermi LAT Collaboration 2009; Zdziarski et al. 2018; we adopt the hypothesis that the Cyg X-3 compact object is a black hole).

For Cyg X-1, gamma-ray observations place the high/soft-state photon energy cut-off at 40–80 MeV (Zdziarski et al. 2017). This is fairly consistent with absorption of gamma-rays propagating through the quasi-thermal radiation field of the disc. Naively, one expects absorption to kick in at around $\sim (m_ec^2)^2/\epsilon_{disc} \sim 300$ MeV. However, the quasi-thermal spectrum is not monochromatic, but falls off smoothly at energies beyond its 1-keV peak. In fact, the disc dominates the observed spectrum at up to 3 or 4 keV (e.g. McConnell et al. 2002), which brings the absorption cut-off into the observed range of 40–80 MeV. Note that the pair-production optical depth $\tau_{\gamma\gamma,disc}$ (equation 127) decreases from $\sim 10^4$ at photon energies beyond the spectral peak, but is still much larger than unity at 3−4 keV. Alternatively, the photon energy cut-off could correspond to the cut-off in the distribution of radiating particles.

Unlike Cyg X-1, Cyg X-3 exhibits gamma-ray emission far above its expected absorption cut-off. However, these high-energy gamma-rays are thought to come from Comptonization of the companion star's radiation by jet electrons at much larger distances (Dubus, Cerutti & Henri 2010; Zdziarski et al. 2018). Thus, gamma-ray observations of Cyg X-3 do not provide a strong constraint on the cut-off in the coronal emission, which may be similar to Cyg X-1 and also consistent with pair absorption.

As for their hard X-ray spectra, Cyg X-1 and Cyg X-3 both display high/soft states with power-law hard X-ray photon number indices $\Gamma_{X,HS} \simeq 2.5$ (McConnell et al. 2002; Zdziarski & Gierliński 2004; Szostek, Zdziarski & McCollough 2008; here 'HS' stands for 'high/soft'). In addition, as detailed by Remillard & McClintock (2006), many black hole X-ray binaries are observed with photon indices clustered around this value in their high/soft states (and also

in 'steep power-law' states where the observed thermal and non-thermal luminosities are comparable). As discussed for FSRQ jets (Section 5.2.4), the relatively steep index $\Gamma_{X,HS} \simeq 2.5$ (corresponding to a reconnection-powered injected electron energy distribution power-law index $p \gtrsim 3$; Section 5.2.4) could be associated with swing cycles in pair-regulated Klein–Nishina reconnection.

Like FSRQs, accreting black holes also exhibit rapid flares. These are observed in the X-rays and have been witnessed on time-scales as short as milliseconds (Gierliński & Zdziarski 2003) – close to the light crossing time of an orbit at $r = 10r_g$ from a black hole with mass $M_{BH} = 10\,M_\odot$. Kinetic beaming in the context of Klein–Nishina reconnection may help explain such rapid variability by delivering observed time-scales as short as 0.1L/c (Cerutti et al. 2013, 2014a, b). However, this has not been explored in detail as in the case of rapid TeV FSRQ flares and requires further attention.

In summary, pair-regulated Klein–Nishina reconnection in the ADCe of black hole X-ray binaries may explain the observed steep spectra from these objects, including the gamma-ray cut-off (observed in Cyg X-1, Zdziarski et al. 2017, though this may also result from the cut-off in the distribution of radiating particles). Kinetic beaming facilitated by Klein–Nishina reconnection may additionally explain short flaring time-scales, but this requires much more thorough investigation.

6 CONCLUSIONS

Collisionless astrophysical reconnection is often radiative, with radiative processes not only producing the observed light, but also coupling to the reconnection dynamics. Many interesting systems realize such a regime where an intense but soft radiation field bathes the reconnection region, and reconnection-energized particles Comptonize the ambient photons to observed X-ray and gamma-ray energies.

Sometimes, IC losses proceed purely in the Thomson limit and the Comptonized photons freely stream out of the system – a regime we discuss in Section 2. Even then, however, radiation is far from passive. The emitting particles may still lose a significant amount of their energy on sub-dynamical time-scales, and this has the potential to modify both reconnection itself and the qualitative features of the received light (e.g. Uzdensky 2016; Beloborodov 2017; Werner et al. 2019; Mehlhaff et al. 2020; Sironi & Beloborodov 2020).

However, there are some astrophysical systems – such as FSRQ blazar jets and the coronae of accreting black holes – where the physical picture is even richer. In these cases, Comptonization of ambient photons can enter the Klein–Nishina regime, and this both qualitatively changes the physics of radiative cooling and opens up an entirely new channel for radiative feedback on reconnection: pair production. By incorporating Klein–Nishina and pair-production physics into the basic conceptual framework of radiative reconnection (Section 3), we find a fundamentally new, self-regulated, Klein–Nishina reconnection regime (Section 4) that stems from a negative feedback loop (Fig. 1). In this loop, gamma-ray photons Comptonized in the reconnection layer propagate into the upstream plasma where they produce hot pairs by colliding with soft (unscattered) ambient photons, thus lowering the effective upstream magnetization. This inhibits further particle acceleration in the reconnection layer, closing the feedback loop.

The pair-regulation mechanism gives rise to an effective fixed-point upstream magnetization σ_h that is *universal* – independent of the initial value $\sigma_{h,0}$. However, the fixed point is not necessarily reached by the system. If pair feedback is efficient enough, the system overshoots its natural solution, σ_h, by a wide margin and enters a

limit (or 'swing') cycle, constantly oscillating between a high- and a low-magnetization state.

In Section 4, we also analyse the number of particles produced in the upstream region. A pair cascade, in which the population in each subsequent newborn generation grows exponentially, is not generally expected (except for unrealistically optimal reconnection-driven NTPA). Even in the presence of a cascade, there is a wide parameter range where the total created pair count is dwarfed by that of the pairs originally present. Thus, remarkably, though the newborn pairs are so hot that they can dominate the upstream energy budget, they are often also so tenuous that they contribute negligibly to the upstream lepton density. This feature, in particular, distinguishes Klein–Nishina radiative reconnection from other pair-regulated regimes (Lyubarskii 1996; Hakobyan et al. 2019).

Finally, in Section 5, we discuss the observable aspects of pair-regulated Klein–Nishina reconnection. We expect that emission from the layer (i.e. the region threaded with reconnected flux) dominates that from the newborn component of the upstream plasma (Section 5.1.1). We also (Section 5.1.2) identify several reasons to expect that Klein–Nishina physics may promote kinetic beaming, which is an important mechanism for facilitating rapid flaring variability in reconnection (Cerutti et al. 2012b, 2013, 2014a, b; Nalewajko et al. 2012; Mehlhaff et al. 2020). Furthermore, we explicitly estimate the parameters (i.e. the energy scales introduced in Sections 2 and 3) governing the radiative regime of reconnection in FSRQ blazar jets (Sections 5.2.1–5.2.4) and ADCe of black hole X-ray binaries (Sections 5.2.5 and 5.2.6). Reconnection in both types of objects is very likely to be strongly impacted by Klein–Nishina physics, and may even enter the pair-regulated regime modelled in this paper. In fact, the generally steep non-thermal spectra produced by FSRQ jets and black hole ADCe appear consistent with strong pair loading in our model, and the observed instances of rapid variability could be caused by the kinetic beaming mechanism. Pair-regulated reconnection may further provide a source for *in situ* pair production in FSRQ jets, where the lepton content is difficult to constrain observationally.

However, many of our observational remarks, as well as the basic physical features of our model, require additional testing through first-principles radiative PIC simulations, which can address some of the following key physical and observational questions:

(i) How robust is the pair feedback mechanism? In particular,

    (a) Is the universal magnetization solution ever realized?

    (b) Can an order-unity energy recapture efficiency ξ be achieved?

    (c) Is the feedback strong enough to induce late-time limit cycles that jump between high- and low-magnetization states?

(ii) How many pairs are produced? Are they indeed few compared to the number of background particles, and does their number density scale with the reconnection parameters as predicted by our model?

(iii) Are the emitted spectra produced by reconnection in the steady state (or in a limit cycle) consistent with our predictions? Are they consistent with observations?

(iv) Can kinetic beaming operate in Klein–Nishina reconnection as we argue? If so, is beaming more or less prominent than in Thomson radiative reconnection (as diagnosed by Mehlhaff et al. 2020)?

(v) To what extent does the physical picture change in 3D, in the presence of ions, with a finite guide field, and (especially) with finite synchrotron losses?

In addition to addressing these immediate follow-up questions, it would be interesting to generalize our theory in several additional ways. For example, one might consider a broader (e.g. power-law) ambient photon spectrum (here we only treat monochromatic ambient seed photons). Beyond shedding light on a broader class of astrophysical systems, such an exercise steps towards a theory of synchrotron self-Compton reconnection, where the dominant population of seed photons for Comptonization is the synchrotron photons produced inside the reconnection system. This, in turn, would be very important for understanding reconnection-powered emission from BL Lac blazars (Madejski & Sikora 2016). As a completely separate extension of this work, kinetic simulations mimicking global jet-like geometries (e.g. Alves, Zrake & Fiuza 2018; Sironi et al. 2021) and also incorporating Klein–Nishina and pair-production physics would determine whether and how this type of reconnection occurs in an explicitly global context. The same could be said for accretion-like numerical set-ups (e.g. Crinquand et al. 2021).

Overall, this study poses a rich set of important questions for future inquiry. In the short term, direct numerical tests of our model will facilitate more straightforward translation to observations. More remotely, global simulations may reveal explicitly how reconnection as modelled here might occur in nature. Finally, the novel radiative physics treated in this work (but see also Lyubarskii 1996; Beloborodov 2017; Hakobyan et al. 2019; Schoeffler et al. 2019; Werner et al. 2019; Sironi & Beloborodov 2020) may help to pave the way to theories of increasingly realistic reconnection regimes with applications to the high-energy Universe.

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

REFERENCES

Abeysekara A. U. et al., 2015, ApJ, 815, L22
Aharonian F. A., Atoyan A. M., 1981, Ap&SS, 79, 321
Aharonian F. A., Belyanin A. A., Derishev E. V., Kocharovsky V. V., Kocharovsky V. V., 2002, Phys. Rev. D, 66, 023005
Ahnen M. L. et al., 2015, ApJ, 815, L23
Aleksić J. et al., 2011, ApJ, 730, L8
Alves E. P., Zrake J., Fiuza F., 2018, Phys. Rev. Lett., 121, 245101
Ball D., Sironi L., Özel F., 2018, ApJ, 862, 80
Begelman M. C., Sikora M., 1987, ApJ, 322, 650
Begelman M. C., Blandford R. D., Rees M. J., 1984, Rev. Mod. Phys., 56, 255
Begelman M. C., Fabian A. C., Rees M. J., 2008, MNRAS, 384, L19
Beloborodov A. M., 2017, ApJ, 850, 141
Beskin V. S., Istomin Y. N., Parev V. I., 1992, Sov. Astron., 36, 642
Blackman E. G., Field G. B., 1994, Phys. Rev. Lett., 72, 494
Blandford R. D., Znajek R. L., 1977, MNRAS, 179, 433
Bloom S. D., Marscher A. P., 1996, ApJ, 461, 657
Blumenthal G. R., Gould R. J., 1970, Rev. Mod. Phys., 42, 237
Boldyrev S., Loureiro N. F., 2020, J. Geophys. Res. (Space Phys.), 125, e28185
Böttcher M., Reimer A., Sweeney K., Prakash A., 2013, ApJ, 768, 54
Cerutti B., Philippov A. A., 2017, A&A, 607, A134
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., Spitkovsky A., 2016, MNRAS, 457, 2401

Cerutti B., Philippov A. A., Dubus G., 2020, A&A, 642, A204

Chen A. Y., Yuan Y., Yang H., 2018, ApJ, 863, L31

Christie I. M., Petropoulou M., Sironi L., Giannios D., 2019, MNRAS, 482, 65

Comisso L., Sironi L., 2018, Phys. Rev. Lett., 121, 255101

Comisso L., Sironi L., 2019, ApJ, 886, 122

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

Dahlin J. T., Drake J. F., Swisdak M., 2014, Phys. Plasmas, 21, 092304

Dermer C. D., Schlickeiser R., 2002, ApJ, 575, 667

Di Matteo T., 1998, MNRAS, 299, L15

Drake J. F., Swisdak M., Che H., Shay M. A., 2006, Nature, 443, 553

Dubus G., Cerutti B., Henri G., 2010, MNRAS, 404, L55

Eardley D. M., Lightman A. P., 1976, Nature, 262, 196

Fermi LAT Collaboration, 2009, Science, 326, 1512

Ford A. L., Keenan B. D., Medvedev M. V., 2018, Phys. Rev. D, 98, 063016

Fossati G., Maraschi L., Celotti A., Comastri A., Ghisellini G., 1998, MNRAS, 299, 433

Galeev A. A., Rosner R., Vaiana G. S., 1979, ApJ, 229, 318

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., Tavecchio F., Chiaberge M., 2005, A&A, 432, 401

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

Gierliński M., Zdziarski A. A., 2003, MNRAS, 343, L84

Gierliński M., Zdziarski A. A., Poutanen J., Coppi P. S., Ebisawa K., Johnson W. N., 1999, MNRAS, 309, 496

Goodman J., Uzdensky D., 2008, ApJ, 688, 555

Gould R. J., Schréder G. P., 1967, Phys. Rev., 155, 1404

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. et al., 2016, ApJ, 818, L9

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

Hakobyan H., Philippov A., Spitkovsky A., 2019, ApJ, 877, 53

Hakobyan H., Petropoulou M., Spitkovsky A., Sironi L., 2021, ApJ, 912, 48

H. E. S. S. Collaboration, 2013, A&A, 554, A107

Hillas A. M., 1984, ARA&A, 22, 425

Jaroschek C. H., Hoshino M., 2009, Phys. Rev. Lett., 103, 075002

Ji H., Daughton W., 2011, Phys. Plasmas, 18, 111207

Jones F. C., 1968, Phys. Rev., 167, 1159

Kagan D., Nakar E., Piran T., 2018, MNRAS, 476, 3902

Li T.-Y., Yorke J. A., 2004, Period Three Implies Chaos. Springer, New York, NY, p. 77

Liang E. P. T., Price R. H., 1977, ApJ, 218, 247

Loureiro N. F., Boldyrev S., 2020, ApJ, 890, 55

Lyubarskii Y. E., 1996, A&A, 311, 172

Lyubarsky Y. E., 2005, MNRAS, 358, 113

Lyutikov M., Uzdensky D., 2003, ApJ, 589, 893

McConnell M. L. et al., 2002, ApJ, 572, 984

McKinney J. C., Uzdensky D. A., 2012, MNRAS, 419, 573

Madejski G. G., Sikora M., 2016, ARA&A, 54, 725

MAGIC Collaboration, 2008, Science, 320, 1752

Malmrose M. P., Marscher A. P., Jorstad S. G., Nikutta R., Elitzur M., 2011, ApJ, 732, 116

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

Melia F., Königl A., 1989, ApJ, 340, 162

Melzani M., Walder R., Folini D., Winisdoerffer C., Favre J. M., 2014, A&A, 570, A111

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., Begelman M. C., Sikora M., 2014, ApJ, 789, 161

Nalewajko K., Yuan Y., Chruślińska M., 2018, J. Plasma Phys., 84, 755840301

Nättilä J., Beloborodov A. M., 2020, preprint (arXiv:2012.03043)

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

Ortuño-Macías J., Nalewajko K., 2020, MNRAS, 497, 1365

Parfrey K., Philippov A., Cerutti B., 2019, Phys. Rev. Lett., 122, 035101

Pétri J., 2012, MNRAS, 424, 2023

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

Philippov A. A., Spitkovsky A., Cerutti B., 2015, ApJ, 801, L19

Philippov A., Uzdensky D. A., Spitkovsky A., Cerutti B., 2019, ApJ, 876, L6

Phinney E. S., 1982, MNRAS, 198, 1109

Poutanen J., Krolik J. H., Ryde F., 1997, MNRAS, 292, L21

Pozdnyakov L. A., Sobol I. M., Syunyaev R. A., 1983, Astrophys. Space Phys. Res., 2, 189

Pushkarev A. B., Kovalev Y. Y., Lister M. L., Savolainen T., 2009, A&A, 507, L33

Remillard R. A., McClintock J. E., 2006, ARA&A, 44, 49

Ripperda B., Bacchini F., Philippov A. A., 2020, ApJ, 900, 100

Rybicki G. B., Lightman A. P., 1979, Radiative Processes in Astrophysics. Wiley, New York

Schoeffler K. M., Grismayer T., Uzdensky D., Fonseca R. A., Silva L. O., 2019, ApJ, 870, 49

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shapiro S. L., Lightman A. P., Eardley D. M., 1976, ApJ, 204, 187

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

Sikora M., Rutkowski M., Begelman M. C., 2016, MNRAS, 457, 1352

Sironi L., Beloborodov A. M., 2020, ApJ, 899, 52

Sironi L., Spitkovsky A., 2014, ApJ, 783, L21

Sironi L., Petropoulou M., Giannios D., 2015, MNRAS, 450, 183

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). Proceedings of Science, Italy, p. 825

Sobacchi E., Nättilä J., Sironi L., 2021, MNRAS, 503, 688

Szostek A., Zdziarski A. A., McCollough M. L., 2008, MNRAS, 388, 1001

Tavani M. et al., 2009, Nature, 462, 620

Tavecchio F., 2017, in Aharonian F. A., Hofmann W., Rieger F. M., eds, AIP Conf. Proc. Vol. 1792, Gamma Rays from Blazars. Am Inst. Phys., New York, p. 020007

Tavecchio F., Ghisellini G., 2008, MNRAS, 386, 945

Tavecchio F., Ghisellini G., 2016, MNRAS, 456, 2374

Tavecchio F., Becerra-Gonzalez J., Ghisellini G., Stamerra A., Bonnoli G., Foschini L., Maraschi L., 2011, A&A, 534, A86

Uzdensky D. A., 2011, Space Sci. Rev., 160, 45

Uzdensky D. A., 2016, Radiative Magnetic Reconnection in Astrophysics. Springer International Publishing, Cham, p. 473

Uzdensky D. A., 2020, preprint (arXiv:2007.09533)

Uzdensky D. A., Goodman J., 2008, ApJ, 682, 608

Uzdensky D. A., McKinney J. C., 2011, Phys. Plasmas, 18, 042105

Uzdensky D. A., Spitkovsky A., 2014, ApJ, 780, 3

Uzdensky D. A., Loureiro N. F., Schekochihin A. A., 2010, Phys. Rev. Lett., 105, 235002

Uzdensky D. A., Cerutti B., Begelman M. C., 2011, ApJ, 737, L40

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

Wong G. N., Ryan B. R., Gammie C. F., 2021, ApJ, 907, 73

Yuan Y., Nalewajko K., Zrake J., East W. E., Blandford R. D., 2016, ApJ, 828, 92

Zdziarski A. A., 1989, ApJ, 342, 1108

Zdziarski A. A., Gierliński M., 2004, Prog. Theor. Phys. Suppl., 155, 99

Zdziarski A. A., Malyshev D., Chernyakova M., Pooley G. G., 2017, MNRAS, 471, 3657

Zdziarski A. A. et al., 2018, MNRAS, 479, 4399

Zhdankin V., Uzdensky D. A., Perez J. C., Boldyrev S., 2013, ApJ, 771, 124

Zhdankin V., Uzdensky D. A., Werner G. R., Begelman M. C., 2020, MNRAS, 493, 603

APPENDIX A: THE KLEIN–NISHINA SCATTERING POWER AND RATE

Here, we sketch the derivations of the functions $f_{KN}(q)$ and $g_{KN}(q)$ as defined in equations (15) and (19). Jones (1968) and Blumenthal & Gould (1970) report the Klein–Nishina scattering kernel

$$\frac{dN}{dt\,dr}\,dr = c\sigma_{T}\frac{U_{ph}}{\epsilon_{ph}}$$
$$\times \frac{3}{(1+qr)^2}\left[2r\ln r + (1+2r)(1-r) + \frac{1}{2}\frac{(qr)^2}{1+qr}(1-r)\right]dr$$
$$\equiv c\sigma_{T}\frac{U_{ph}}{\epsilon_{ph}}K(r,q)dr . \tag{A1}$$

This is the number of photons scattered per unit time by a particle with Klein–Nishina parameter $q = \gamma/\gamma_{KN} = 4\gamma\epsilon_{ph}/m_ec^2$ to final photon energies between r and $r + dr$. The parameter r is not actually equal to the scattered photon energy ϵ_1 but is defined in terms of it through

$$r \equiv \frac{\epsilon_1/\gamma m_e c^2}{q\left(1 - \epsilon_1/\gamma m_e c^2\right)} . \tag{A2}$$

Equivalently,

$$\epsilon_1 = \gamma m_e c^2 \frac{qr}{1+qr} . \tag{A3}$$

It is convenient to take integrals over r as a proxy for ϵ_1 because, as ϵ_1 spans its kinematically allowed range $[\epsilon_{ph}, \gamma m_e c^2 q/(1+q)]$, r runs approximately from 0 to 1 (Jones 1968; Blumenthal & Gould 1970).

For completeness, we mention the validity conditions for the kernel (A1): $\gamma m_e c^2/\epsilon_{ph} = 4\gamma_{KN}\gamma \gg 1$ and $\gamma \gg 1$ (Jones 1968). The first condition is always satisfied for $\gamma_{KN} \gg 1$, which is true of every astrophysical system to which we apply our results (and also of many others). The relativistic requirement $\gamma \gg 1$ is then taken care of because the general Klein–Nishina expressions are only needed for $\gamma \gg \gamma_{KN} \gg 1$. At non-relativistic energies $\gamma \simeq 1$, the Thomson prefactors in equations (14) and (18) are all that remain [$f_{KN}(q \ll 1) \rightarrow 1$ and $g_{KN}(q \ll 1) \rightarrow 1$], and these are non-relativistically correct. Thus, for $\gamma_{KN} \gg 1$, equations (14) and (18) are correct even in the non-relativistic case.

To obtain the rates at which a scattering particle loses energy (equation 14) and encounters soft seed photons (equation 18) requires

integrating over ϵ_1 – or, equivalently, over r – while keeping γ, ϵ_{ph}, and, hence, $q = \gamma/\gamma_{KN}$ fixed. One obtains

$$\begin{bmatrix} R_{IC}(\gamma) \\ P_{IC}(\gamma) \end{bmatrix} = c\sigma_T \frac{U_{ph}}{\epsilon_{ph}} \int_0^1 dr\, K(r,q) \begin{bmatrix} 1 \\ \epsilon_1(r) \end{bmatrix}$$
$$= \begin{bmatrix} c\sigma_T U_{ph}/\epsilon_{ph} \\ (4/3)c\sigma_T\gamma^2 U_{ph} \end{bmatrix} \int_0^1 dr\, K(r,q) \begin{bmatrix} 1 \\ 3r/(1+qr) \end{bmatrix} . \tag{A4}$$

Then, using $\beta^2 = 1 - 1/\gamma^2 \simeq 1$ to write $P_T(\gamma) \simeq (4/3)c\sigma_T\gamma^2 U_{ph}$, one can read off $g_{KN}(q)$ and $f_{KN}(q)$ from equation (A4):

$$g_{KN}(q) = \int_0^1 dr\, K(r,q) \tag{A5}$$

and

$$f_{KN}(q) = 3\int_0^1 dr\, K(r,q)\frac{r}{1+qr} . \tag{A6}$$

The integrals can be evaluated to yield equations (15) and (19). Should the reader wish to verify by explicit computation, we find that the identity

$$\mathrm{Li}_2\left(\frac{1}{1+q}\right) = \mathrm{Li}_2(-q) + \frac{\pi^2}{6} + \frac{1}{2}\log^2(q) - \frac{1}{2}\log^2\left(\frac{q}{1+q}\right) \tag{A7}$$

is useful.

APPENDIX B: THE ENERGY RETENTION FACTOR f_{NOCOOL}

Recall that in Section 4.1, we left ξ as a free parameter. This stemmed chiefly from our uncertainty regarding the fraction of energy f_{nocool} retained by the fresh pair plasma as it travels towards the reconnection layer. In order to demonstrate the effect of this parameter without knowing it precisely, we entertained a class of models where ξ (specifically f_{nocool}) was constant: independent of NTPA in the layer [i.e. of $p(\sigma_h)$ and z]. However, in principle, we can explicitly compute the distributions $N_{\gamma\gamma}^{(n)}(\gamma)$ of newborn pairs (by straightforwardly extending the analysis of Section 4.2), and, from them, calculate the energy carried by fresh plasma reaching the layer. This can then be compared to the energy *deposited* into the upstream region to yield f_{nocool}.

In this section, we adopt this strategy to calculate $\xi = f_{nocool}$. We find that, like $\mathcal{F}$, ξ is a function of $p(\sigma_h)$ and z: $\xi = \Xi[p(\sigma_h), z]$. Furthermore, the basic intuition gleaned in Section 4.1 remains intact: extremely efficient energy delivery to the layer, $\xi \sim 1$, may still push the system into a two-state swing cycle, and radiative feedback on NTPA still makes the system more prone to these oscillations. Through the present analysis, we merely gain a more precise notion of the values of physical parameters – particularly the cut-off z in the distribution of layer particles – required to initiate swing cycles versus those that cause rapid progression towards the fixed point σ_h.

Our first step is to write down a quantitative expression for f_{nocool}. In order to do this, we define several auxiliary quantities: the nth-generation steady-state particle count

$$N_{\gamma\gamma}^{(n)} \equiv \int N_{\gamma\gamma}^{(n)}(\gamma)d\gamma , \tag{B1}$$

which we distinguish from $N_{\gamma\gamma}^{(n)}(\gamma)$ by omitting the functional argument; the nth-generation particle injection rate

$$Q_{\gamma\gamma}^{(n)} \equiv \int Q_{\gamma\gamma}^{(n)}(\gamma)d\gamma , \tag{B2}$$

also distinguished from $Q_{\gamma\gamma}^{(n)}(\gamma)$ by argument omission; the average particle energy *injected* into the nth generation

$$\bar{\gamma}_Q^{(n)} \equiv \frac{\int \gamma \, Q_{\gamma\gamma}^{(n)}(\gamma) \mathrm{d}\gamma}{Q_{\gamma\gamma}^{(n)}}\,; \tag{B3}$$

and the average particle energy *carried* by the nth generation in its steady state

$$\bar{\gamma}_N^{(n)} \equiv \frac{\int \gamma \, N_{\gamma\gamma}^{(n)}(\gamma) \mathrm{d}\gamma}{N_{\gamma\gamma}^{(n)}}\,. \tag{B4}$$

In terms of these quantities, one has

$$
\begin{aligned}
f_{\text{nocool}} &\equiv \frac{\text{newborn pair energy carried into layer}}{\text{energy injected upstream}} \\
&= \frac{\sum_{n=1}^{N} N_{\gamma\gamma}^{(n)} \bar{\gamma}_N^{(n)}}{\sum_{n=1}^{N} N_{\gamma\gamma}^{(n)} \bar{\gamma}_Q^{(n)}} \\
&= \frac{\sum_{n=1}^{N} N_{\gamma\gamma}^{(n)} \bar{\gamma}_Q^{(n)} f_{\text{nocool}}^{(n)}}{\sum_{n=1}^{N} N_{\gamma\gamma}^{(n)} \bar{\gamma}_Q^{(n)}}\,,
\end{aligned} \tag{B5}
$$

where

$$f_{\text{nocool}}^{(n)} \equiv \frac{N_{\gamma\gamma}^{(n)} \bar{\gamma}_N^{(n)}}{N_{\gamma\gamma}^{(n)} \bar{\gamma}_Q^{(n)}} = \frac{\bar{\gamma}_N^{(n)}}{\bar{\gamma}_Q^{(n)}} \tag{B6}$$

is the average fractional energy retained by the nth generation.

Before we calculate f_{nocool} in detail using equation (B5), let us briefly argue why, as claimed in Section 4.1, f_{nocool} should be physically confined to the range [3/400, 1]. The upper bound is trivial: f_{nocool} cannot exceed one by definition. Let us now see how the lower bound arises.

In the case where f_{nocool} is as small as possible, all particles born into the upstream region are quite energetic, with $\gamma_{\text{KN}} \ll \bar{\gamma}_Q^{(n)} \lesssim \bar{\gamma}_{\text{cool},2}$, but they cool quickly – until their cooling times match their readvection time t_{ra}, and hence until their Lorentz factors equal $\bar{\gamma}_{\text{cool},1}$ (equation 59). However, when these particles' energies exceed γ_{pp}, none of their emitted photons escape the system. Rather, the power $Q_{\gamma\gamma}^{(1)} \bar{\gamma}_Q^{(1)} m_e c^2$ injected into the upstream region losslessly converts to secondary pairs, and therefore matches the steady-state energy flux $Q' \gamma_{\text{pp}} m_e c^2$ into the particle energy bin $\gamma < \gamma_{\text{pp}} < \gamma + \mathrm{d}\gamma$. Here, Q' is the (generationally summed) particle flux into the same bin. Only once particles cool past γ_{pp} does their emission leak out of the system. Thus, if the injected particles have high initial energies but all cool down to $\bar{\gamma}_{\text{cool},1}$, f_{nocool} reaches the (lowest possible) value

$$
\begin{aligned}
f_{\text{nocool,min}} &\equiv 1 - \frac{\text{power radiated by upstream particles}}{\text{power injected upstream}} \\
&= 1 - \frac{Q'(\gamma_{\text{pp}} - \bar{\gamma}_{\text{cool},1})}{Q_{\gamma\gamma}^{(1)} \bar{\gamma}_Q^{(n)}} \\
&= 1 - \frac{\gamma_{\text{pp}} - \bar{\gamma}_{\text{cool},1}}{\gamma_{\text{pp}}} \\
&= \frac{\bar{\gamma}_{\text{cool},1}}{\gamma_{\text{pp}}} = \frac{1}{8}\frac{3}{50} = \frac{3}{400}\,.
\end{aligned} \tag{B7}
$$

The reason that (as we show quantitatively in this section) $f_{\text{nocool, min}} = 3/400$ is an overly pessimistic estimate for f_{nocool} is that, in reality, upstream particles are constantly being replenished at high energies even as they rapidly cool towards $\bar{\gamma}_{\text{cool},1}$. This results in a pronounced high-energy tail of hot upstream particles entering the reconnection layer, greatly enhancing f_{nocool}, even, in some cases, to order unity.

B1 The steady-state pair distributions $N_{\gamma\gamma}^{(n)}(\gamma)$

Evaluating equation (B5) for f_{nocool} explicitly requires the steady-state pair distributions $N_{\gamma\gamma}^{(n)}(\gamma)$. We now retrieve these distributions by extending the analysis of Section 4.2. Rearranging equation (66), one may write

$$N_{\gamma\gamma}^{(n)}(\gamma) = \frac{1}{8} t_{\text{cool,IC}}(\gamma) Q_{\gamma\gamma}^{(n+1)}(\gamma/4)\,. \tag{B8}$$

Plugging in $t_{\text{cool,IC}}(\gamma) \propto [\gamma f_{\text{KN}}(\gamma/\gamma_{\text{KN}})]^{-1}$ and using the approximate expression $f_{\text{KN}}(q) \simeq (1 + q)^{-1.5}$ from equation (17), one sees that $N_{\gamma\gamma}^{(n)}(\gamma) \propto \gamma^{-\Gamma+0.5}$ provided $Q_{\gamma\gamma}^{(n)} \propto \gamma^{-\Gamma}$ (i.e. as in 57) and $\gamma \gg \gamma_{\text{KN}}$. The latter condition is satisfied because the deep Klein–Nishina regime is assumed in writing equations (63) and (64), from which (B8) follows.

However, this is not the full story for $N_{\gamma\gamma}^{(n)}(\gamma)$. Even though $Q_{\gamma\gamma}^{(n)}(\gamma)$ is zero when $\gamma < \gamma_{\text{min}} \equiv 2\gamma_{\text{KN}}$, which comes from the fact that no pairs are injected at energies below γ_{min}, $N_{\gamma\gamma}^{(n)}(\gamma)$ is *not* zero at these low energies. Instead, particles in every generation cool continuously – in the Thomson regime – once they reach energies $\sim\gamma_{\text{KN}} = \gamma_{\text{min}}/2$, populating a low-energy component of each distribution $N_{\gamma\gamma}^{(n)}(\gamma)$.

To model this situation, we assume that, in each generation, a constant flux of particles leaks from above to below γ_{KN} in energy space [through the advective term in equation (63), which remains nonzero in the Thomson regime even though $Q_{\gamma\gamma}^{(n)}(\gamma)$ vanishes]. Once there, the rate of change of each particle's energy is approximately Thomson: $-\dot{\gamma} \simeq -\dot{\gamma}_{\text{T}} \equiv \gamma/t_{\text{cool,T}}(\gamma)$. To determine $N_{\gamma\gamma}^{(n)}(\gamma)$ in the Thomson limit $\gamma < \gamma_{\text{KN}}$ then requires solving a simplified version of equation (63),

$$\frac{\partial}{\partial t} N_{\gamma\gamma}^{(n)}(\gamma, t) + \frac{\partial}{\partial \gamma}\left(\dot{\gamma}_{\text{T}} N_{\gamma\gamma}^{(n)}(\gamma, t)\right) = Q_{\gamma\gamma}^{(n)} \delta(\gamma - \gamma_{\text{KN}})\,, \tag{B9}$$

where the source term $Q_{\gamma\gamma}^{(n)} \delta(\gamma - \gamma_{\text{KN}}) = \delta(\gamma - \gamma_{\text{KN}}) \int Q_{\gamma\gamma}^{(n)}(\gamma) \mathrm{d}\gamma$ gives the flux into γ_{KN} of particles from higher energies.[8] The solution to (B9) can be obtained exactly, and the steps are detailed in Appendix E. The result is a low-energy distribution of particles arriving at the layer in each generation of

$$N_{\gamma\gamma}^{(n)}(\gamma \lesssim \gamma_{\text{KN}}) = Q_{\gamma\gamma}^{(n)} \gamma_{\text{cool}} \frac{L}{c} \begin{cases} \gamma^{-2} & \gamma \geq \bar{\gamma}_{\text{cool},1} \\ 0 & \gamma < \bar{\gamma}_{\text{cool},1} \end{cases}\,. \tag{B10}$$

Equivalently, one can use $\gamma_{\text{cool}} = (10/\tau_{\gamma\gamma})\bar{\gamma}_{\text{cool},1} = (10/\tau_{\gamma\gamma})(3/50)\gamma_{\text{KN}}$ to express this as

$$N_{\gamma\gamma}^{(n)}(\gamma \lesssim \gamma_{\text{KN}}) = Q_{\gamma\gamma}^{(n)} \frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c} \begin{cases} \gamma_{\text{KN}}/\gamma^2 & \gamma \geq \bar{\gamma}_{\text{cool},1} \\ 0 & \gamma < \bar{\gamma}_{\text{cool},1} \end{cases}\,, \tag{B11}$$

which will be a slightly more useful form later. For reference, the average energy of this distribution is

$$
\begin{aligned}
\frac{\int_1^{\gamma_{\text{min}}} \gamma \, N_{\gamma\gamma}^{(n)}(\gamma)\mathrm{d}\gamma}{\int_1^{\gamma_{\text{min}}} N_{\gamma\gamma}^{(n)}(\gamma)\mathrm{d}\gamma} &= \bar{\gamma}_{\text{cool},1} \ln\left(\frac{2\gamma_{\text{KN}}}{\bar{\gamma}_{\text{cool},1}}\right) \\
&\simeq 3.5\bar{\gamma}_{\text{cool},1}\,,
\end{aligned} \tag{B12}
$$

where the integrals are taken through $\gamma_{\text{min}} = 2\gamma_{\text{KN}}$ because, as detailed below, this is where we match the Thomson solution (B11) to the one we will obtain in the deep Klein–Nishina regime.

Next, we determine the high-energy ($\gamma > \gamma_{\text{KN}}$) part of the steady-

---

[8] Strictly speaking, the first generation of pairs $Q_{\gamma\gamma}^{(1)}(\gamma)$ may have particles at energies between $\bar{\gamma}_{\text{cool},2}$ and $\gamma_2^{(1)} = \gamma_2/4$ if $\gamma_2 > 4\bar{\gamma}_{\text{cool},2}$. In that case, taking $Q_{\gamma\gamma}^{(n)} = \int Q_{\gamma\gamma}^{(n)}(\gamma)\mathrm{d}\gamma$, and not cutting off the integral at $\bar{\gamma}_{\text{cool},2}$, overestimates the number of particles in the low-energy Thomson regime, artificially reducing our estimate of $\bar{\gamma}_N^{(n)}$ and, hence, of $f_{\text{nocool}}^{(n)}$.
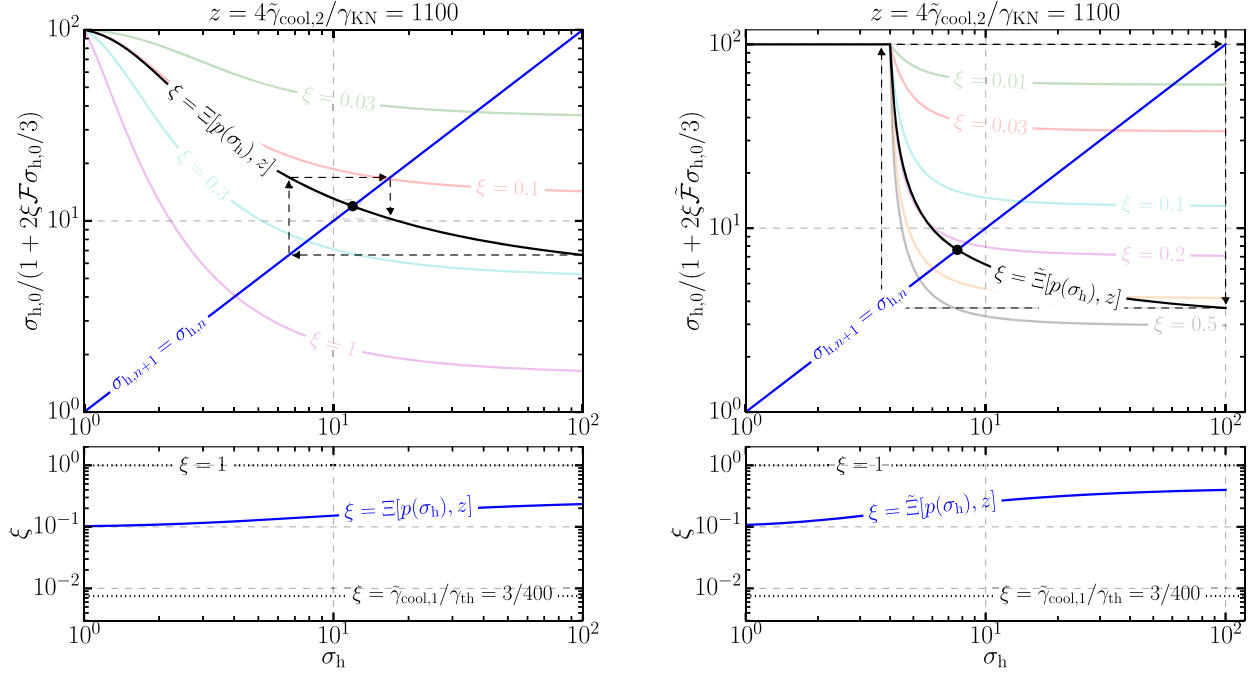
Figure B1. Top subpanels: On the left (right), the function $H(\sigma_h)$ [$\tilde{H}(\sigma_h)$], which includes a calculation of ξ including (excluding) the effect of radiation reaction on the distribution of radiating particles in the reconnection layer, for the case $z = \gamma_2/\gamma_{KN} = 1100$. This function is overlaid on several constant-ξ contours of $h(\sigma_h)$ [$\tilde{h}(\sigma_h)$], which are taken from Fig. 12 (Fig. 15, left-hand panel). When a contour is crossed, it means that the value of $\xi = \Xi$ ($\xi = \tilde{\Xi}$) equals the corresponding contour value. Bottom subpanels: The function Ξ ($\tilde{\Xi}$) plotted explicitly, alongside the minimum and maximum allowed values, $3/400 \sim 0.01$ and 1, respectively. The energy recapture efficiency ξ is always much higher than the minimum, regardless of the assumptions made about radiation reaction, and, if radiation reaction is accounted for, may even approach order unity.

state distributions $N_{\gamma\gamma}^{(n)}(\gamma)$. We note that $Q_{\gamma\gamma}^{(n)}$ can be written (cf. equation 74) as

$$Q_{\gamma\gamma}^{(n)} \simeq B_n \frac{\gamma_{\min}^{-\Gamma+1}}{\Gamma - 1}\left[1 - \left(\frac{\gamma_{\min}}{\gamma_2^{(n)}}\right)^{\Gamma-1}\right] \simeq B_n \frac{\gamma_{\min}^{-\Gamma+1}}{\Gamma - 1}. \quad (B13)$$

Plugging this into the right-hand side of (B8) gives

$$N_{\gamma\gamma}^{(n)}(\gamma \gtrsim \gamma_{KN}) = \frac{1}{8}\left[t_{\text{cool,IC}}(\gamma)\right]\left[Q_{\gamma\gamma}^{(n+1)}(\gamma/4)\right]$$

$$= \frac{1}{8}\left[\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}\frac{1}{qf_{KN}(q)}\right]\left[B_{n+1}\left(\frac{\gamma}{4}\right)^{-\Gamma}\right]$$

$$= \frac{1}{8}\left[\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}\frac{1}{qf_{KN}(q)}\right]\left[\mathcal{A}(\Gamma)B_n 4^\Gamma \gamma^{-\Gamma}\right]$$

$$\simeq \frac{1}{8}\left[\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}\frac{1}{qf_{KN}(q)}\right]\left[\mathcal{A}(\Gamma)Q_{\gamma\gamma}^{(n)}4^\Gamma(\Gamma-1)\gamma_{\min}^{-1}\left(\frac{\gamma}{\gamma_{\min}}\right)^{-\Gamma}\right]$$

$$= \frac{1}{2\gamma_{KN}}Q_{\gamma\gamma}^{(n)}\left[\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}\frac{1}{qf_{KN}(q)}\right]\left(\frac{\gamma}{\gamma_{\min}}\right)^{-\Gamma}. \quad (B14)$$

The third line follows from (72), the fourth line from (B13), and the last line from (73) along with $\gamma_{\min} = 2\gamma_{KN}$. Rigorously matching the low-energy Thomson solution (B11) to the high-energy Klein–Nishina result (B14) requires a detailed analysis of the radiative physics near $\gamma \sim \gamma_{KN}$. We expect this, at most, to modify (B11) and (B14) by order-unity factors near $\gamma \sim \gamma_{KN}$, since the assumptions from which these solutions are derived are fairly robust in their respective limits (far from γ_{KN}). Moreover, the solutions are already of similar scale at a natural matching point, $\gamma = \gamma_{\min} = 2\gamma_{KN}$, with equation (B14) a factor of about $1/f_{KN}(2) \sim 5$ larger than (B11) when both are evaluated at $\gamma_{\min}$. This is independent of Γ. Thus, as a rough estimate, we take the overall distribution of nth-generation

pairs (including both $\gamma > \gamma_{KN}$ and $\gamma < \gamma_{KN}$) to be

$$N_{\gamma\gamma}^{(n)}(\gamma) \simeq Q_{\gamma\gamma}^{(n)}\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}$$

$$\times \begin{cases} \gamma_{\min}/2\gamma^2 & \tilde{\gamma}_{\text{cool,1}} \leq \gamma < \gamma_{\min} \\ [\gamma_{\min}qf_{KN}(q)]^{-1}(\gamma/\gamma_{\min})^{-\Gamma} & \gamma_{\min} \leq \gamma < \gamma_2^{(n)} \\ 0 & \text{otherwise} \end{cases}. \quad (B15)$$

If one leverages the fact that $q > 2$ when $\gamma > \gamma_{\min}$ to approximate $qf_{KN}(q) \simeq q^{-0.5}$, equation (B15) becomes

$$N_{\gamma\gamma}^{(n)}(\gamma) \simeq Q_{\gamma\gamma}^{(n)}\frac{3}{50}\frac{10L}{\tau_{\gamma\gamma}c}$$

$$\times \begin{cases} \gamma_{\min}/2\gamma^2 & \tilde{\gamma}_{\text{cool,1}} \leq \gamma < \gamma_{\min} \\ (\sqrt{2}/\gamma_{\min})(\gamma/\gamma_{\min})^{-\Gamma+0.5} & \gamma_{\min} \leq \gamma < \gamma_2^{(n)} \\ 0 & \text{otherwise} \end{cases}. \quad (B16)$$

B2 Evaluating f_{nocool}

Armed with the pair distributions $N_{\gamma\gamma}^{(n)}(\gamma)$, and knowing, from Section 4.2, the injected distributions $Q_{\gamma\gamma}^{(n)}(\gamma)$, we can now explicitly evaluate equation (B5) to obtain f_{nocool}. Assembling all of the ingredients – equations (57), (B1)–(B4), (B13), and (B16) – we have

$$\frac{\bar{\gamma}_Q^{(n)}}{\gamma_{\text{pp}}} = \frac{1}{4}\frac{\Gamma-1}{\Gamma-2}\left[\frac{1-(z/4^{n+1/2})^{2-\Gamma}}{1-(z/4^{n+1/2})^{1-\Gamma}}\right], \quad (B17)$$

$$N_{\gamma\gamma}^{(n)} = Q_{\gamma\gamma}^{(n)}\frac{10L}{\tau_{\gamma\gamma}c}\left\{1 + \sqrt{2}\frac{3/50}{\Gamma-1.5}\left[1 - \left(\frac{z}{4^{n+1/2}}\right)^{1.5-\Gamma}\right]\right\}, \quad (B18)$$

and

$$\frac{\bar{\gamma}_N^{(n)}}{\gamma_{\rm pp}} = \frac{3}{400} \left\{ 3.5 + \frac{2\sqrt{2}}{2.5 - \Gamma} \left[\left(\frac{z}{4^{n+1/2}} \right)^{2.5-\Gamma} - 1 \right] \right\}$$
$$\times \left\{ 1 + \sqrt{2} \frac{3/50}{\Gamma - 1.5} \left[1 - \left(\frac{z}{4^{n+1/2}} \right)^{1.5-\Gamma} \right] \right\}^{-1} . \qquad (B19)$$

Some special numbers that appear in these expressions are: $3.5 \simeq \ln(2 \times 3/50) = \ln(\gamma_{\rm min}/\tilde{\gamma}_{\rm cool,1})$ in the numerator of (B19) (cf. equation B12); $1/4 \equiv \gamma_{\rm min}/\gamma_{\rm pp}$, the prefactor of (B17); and $3/400 \equiv \tilde{\gamma}_{\rm cool,1}/\gamma_{\rm pp} = f_{\rm nocool,min}$, the prefactor of (B19). In writing these expressions, we assume that the layer particle distribution cuts off at $\gamma_2 \leq 4\tilde{\gamma}_{\rm cool,2}$ and, therefore, that we are justified to replace $\gamma_2^{(n)}/\gamma_{\rm min}$ with $\gamma_2/2\gamma_{\rm KN}4^n = z/4^{n+1/2}$, as we have done.

Some important features and limits of these formulae are as follows. First, noting that $\Gamma \simeq p_{\rm KN} + 1 \simeq p + 0.5$, we see that $\Gamma \in [1.5, 3.5]$. Thus, the second term in the braces of equation (B18) is almost always small, and, generally, $N_{\gamma\gamma}^{(n)} \sim Q_{\gamma\gamma}^{(n)} 10L/\tau_{\gamma\gamma c}$: the number of particles in the steady state is dominated by those in the Thomson regime. Next, when Γ is on the softer side ($\Gamma \gtrsim 2.5$), all expressions become virtually z-independent (unless $z \sim 4^{n+1/2}$). This is because the distributions of pairs are so steep that all quantities are dominated by the low energies [near $\gamma_{\rm min}$ for the $Q_{\gamma\gamma}^{(n)}(\gamma)$'s and near $\tilde{\gamma}_{\rm cool,1}$ for the $N_{\gamma\gamma}^{(n)}(\gamma)$'s]. In this limit, the weights $\bar{\gamma}_Q^{(n)} N_{\gamma\gamma}^{(n)}/\sum_k \bar{\gamma}_Q^{(k)} N_{\gamma\gamma}^{(k)}$ used to average the $f_{\rm nocool}^{(n)}$'s in (B5) exhibit a very simple n-dependence: $\bar{\gamma}_Q^{(n)} N_{\gamma\gamma}^{(n)} \propto Q_{\gamma\gamma}^{(n)} \propto \mathcal{A}(\Gamma)^n \ll 1$. Thus, $f_{\rm nocool}$ is dominated by the first generation, tending to

$$\lim_{\Gamma \to 3.5} f_{\rm nocool} = \lim_{\Gamma \to 3.5} f_{\rm nocool}^{(1)} = \lim_{\Gamma \to 3.5} \frac{\bar{\gamma}_N^{(n)}}{\bar{\gamma}_Q^{(n)}}$$
$$= \frac{3}{400} \frac{3.5 + 2\sqrt{2}}{(3.5 - 1)/4(3.5 - 2)} \simeq 0.1 . \qquad (B20)$$

This is much higher than our pessimistic estimate $f_{\rm nocool,\,min} = 3/400$, and owes to the extended nature of the power-law distribution of particles. Though dominated by the low energies, this distribution still carries an average particle energy that is a factor of several higher than $\tilde{\gamma}_{\rm cool,1}$.

Finally, we note that $\bar{\gamma}_N^{(n)}$ becomes highly z-dependent when $\Gamma < 2.5$, and so does $\bar{\gamma}_Q^{(n)}$ when $\Gamma < 2$. As the result of this, the highest values of z that we study – all the way up to $z = 4\tilde{\gamma}_{\rm cool,2}/\gamma_{\rm KN} \simeq 1100$ – produce $f_{\rm nocool}$ of order unity (between 0.4 and 0.5). Thus, in our quantitative framework, $f_{\rm nocool}$ is rather weakly dependent on the physical parameters z and p, only varying between about 0.1 and 0.5 over the broad parameter space, $(p, z) \in [1, 3] \times [8, 1100]$. However, one should bear in mind that we have made many simplifying assumptions in our model, and so a wider range of $f_{\rm nocool}$ may be possible in reality. Thus we would like to stress the overall qualitative insight – that $f_{\rm nocool}$ is never really too small (always at least several per cent) and may potentially reach order unity – more than our exact quantitative values. So, even though we now repeat our analysis from Section 4.1 using our functional form for $f_{\rm nocool}$ (via those for $\bar{\gamma}_N^{(n)}$, $\bar{\gamma}_Q^{(n)}$ and $N_{\gamma\gamma}^{(n)}$), we aim to stress general features, showing how the (now somewhat more self-consistent) dynamics predicted by our model depend in a very qualitative sense on z and p.

B3 Reconnection dynamics with self-consistent $f_{\rm nocool}$

Putting $f_{\rm noesc} = 1$ as in Section 4.1 (see also Appendix C), the energy recapture efficiency is just $\xi = f_{\rm noesc} f_{\rm nocool} = f_{\rm nocool}$. We denote the (new in this appendix) self-consistent functional dependence of ξ,

through $f_{\rm nocool}$, on $p_{\rm KN}$ and z by writing

$$\xi = \Xi(p_{\rm KN}, z) = f_{\rm nocool} , \qquad (B21)$$

where $f_{\rm nocool}$ is evaluating according to equation (B5) using equations (B13) and (B17)-(B19). Note that Ξ is a function of p only through $p_{\rm KN}$. This is because the power-law scaling of the upstream particle distributions Γ is inherited strictly from the deep Klein–Nishina-regime layer particles: $\Gamma \simeq p_{\rm KN} + 1$. Equation (B21) eliminates ξ as a free parameter from the problem – just like $\mathcal{F}$, it is entirely determined in terms of $p_{\rm KN}$ (and, hence, in terms of $\sigma_{\rm h}$) and $z = \gamma_2/\gamma_{\rm KN}$. This allows us to repeat our analysis from Section 4.1, except using our expression for ξ in (B21) instead of scanning across it as a free parameter. This we do in Figs B1–B3.

Each of these figures presents results from two cases, one where radiation back reaction on the layer distribution of emitting particles is ignored [i.e. where $p_{\rm T} = p_{\rm KN} = p(\sigma_{\rm h})$] and one where it is crudely taken into account [i.e. where $p_{\rm T} = p(\sigma_{\rm h}) + 1$ and $p_{\rm KN} = p(\sigma_{\rm h}) - 0.5$]. In the former case, we generalize equation (51) to

$$\sigma_{{\rm h},n+1} = H(\sigma_{{\rm h},n}) \equiv \frac{\sigma_{\rm h,0}}{1 + 2\, \Xi[p(\sigma_{{\rm h},n}), z] \mathcal{F}[p(\sigma_{{\rm h},n}), z] \sigma_{\rm h,0}/3} , \quad (B22)$$

whereas in the latter we generalize equation (55) to

$$\sigma_{{\rm h},n+1} = \tilde{H}(\sigma_{{\rm h},n}) \equiv \frac{\sigma_{\rm h,0}}{1 + 2\, \tilde{\Xi}[p(\sigma_{{\rm h},n}), z] \tilde{\mathcal{F}}[p(\sigma_{{\rm h},n}), z] \sigma_{\rm h,0}/3} . \quad (B23)$$

Here, we have defined

$$\tilde{\Xi}(p, z) \equiv \Xi(p - 0.5, z) . \qquad (B24)$$

We note that, when evaluating Ξ and $\tilde{\Xi}$ through equation (B5), we replace the symbol $Q_{\gamma\gamma}^{(n)}$ in the expression for $N_{\gamma\gamma}^{(n)}$ (equation B18) with $B_n \gamma_{\rm min}^{-\Gamma+1}[1 - (4^{n+1/2}/z)^{\Gamma-1}]/(\Gamma - 1)$ rather than adopt the cruder approximation $Q_{\gamma\gamma}^{(n)} \simeq B_n \gamma_{\rm min}^{-\Gamma+1}/(\Gamma - 1)$ (see equation B13). This maintains continuity of $f_{\rm nocool}$, ensuring that contributions from each successive generation turn on gradually with z (as they do in reality) rather than discretely. We have checked that this does not introduce significant error into our calculation, even though it is slightly inconsistent with the derivations of equations (B17)–(B19) (particularly one of the steps in equation B14), which assume $Q_{\gamma\gamma}^{(n)} \simeq B_n \gamma_{\rm min}^{-\Gamma+1}/(\Gamma - 1)$.

In Figs B1–B3, we analyse more thoroughly the dependence of the solution $\sigma_{\rm h}$ on z than in Section 4.1. There are two reasons for this. First, eliminating ξ as a free parameter renders z the only independent variable in the problem, and so examining z-dependence is now easier. Secondly, unlike when we treated ξ as a free parameter, Ξ and $\tilde{\Xi}$ both depend on z, so examining z-dependence is now more necessary. We conduct our analysis in the broad range of $z = \gamma_2/\gamma_{\rm KN}$ spanning from the minimum for pair production to occur in the upstream region, $z = 8$, to the maximum such that all newborn particles are injected with Lorentz factors less than $\tilde{\gamma}_{\rm cool,2}$. In the latter case, z is determined by setting the cut-off in the first generation's injected distribution $\gamma_2^{(1)}$ to $\tilde{\gamma}_{\rm cool,2}$, yielding $z = \gamma_2/\gamma_{\rm KN} = (4\gamma_2^{(1)})/\gamma_{\rm KN} = 4\tilde{\gamma}_{\rm cool,2}/\gamma_{\rm KN} \simeq 1100$ (equation 60).

Fig. B1 shows how the solutions $\sigma_{\rm h} = H(\sigma_{\rm h})$ and $\sigma_{\rm h} = \tilde{H}(\sigma_{\rm h})$, where ξ is determined self-consistently, differ from the corresponding solutions $\sigma_{\rm h} = h(\sigma_{\rm h})$ and $\sigma_{\rm h} = \tilde{h}(\sigma_{\rm h})$, when ξ is taken to be constant. This is done for the extreme case $z = 1100$, which is most suitable for comparing with Section 4.1, where z is infinite. Additionally, Fig. B1 also displays the functions Ξ and $\tilde{\Xi}$. One can see that, at small $\sigma_{\rm h}$, Ξ and $\tilde{\Xi}$ are small ($\sim$0.1, as predicted by equation B20). Furthermore, Ξ and $\tilde{\Xi}$ both increase with $\sigma_{\rm h}$. This results from the power-law index $p(\sigma_{\rm h})$ of radiating layer particles becoming harder, which causes $f_{\rm nocool}$ to be dominated
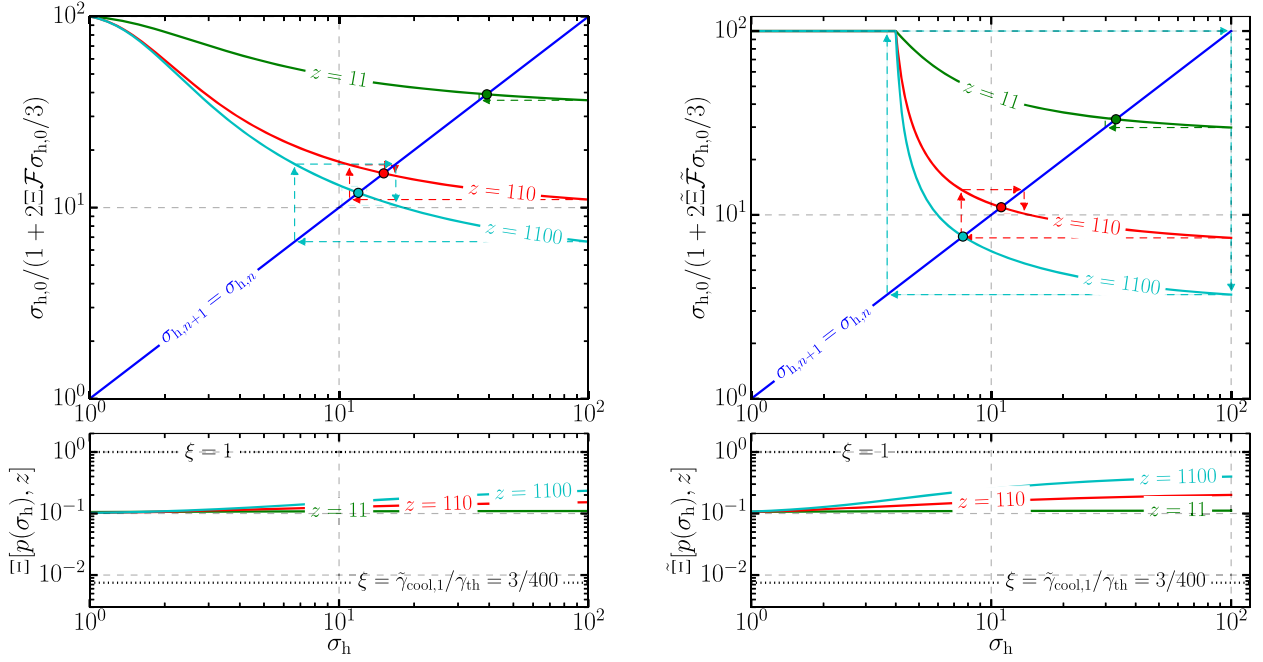
Figure B2. The same as Fig. B1, but without constant-ξ contours of the functions h and $\tilde{h}$. Instead, on the left (right), the functions H ($\tilde{H}$) and Ξ ($\tilde{\Xi}$) are plotted for several different values of z. On both sides, the maximum ξ-value occurs at high-σ_h and is sensitive to z. Both radiation reaction and high-z are required for ξ to reach order unity. In contrast, the low value of ξ, occurring at low-σ_h, is virtually independent of z and is always around $\xi = 0.1 \gg f_{\text{nocool,min}}$. Thus, order-unity efficiency requires both high-z and high-σ_h, but extremely low efficiency ($\xi \sim 10^{-2}$) is averted across the domains of all models.
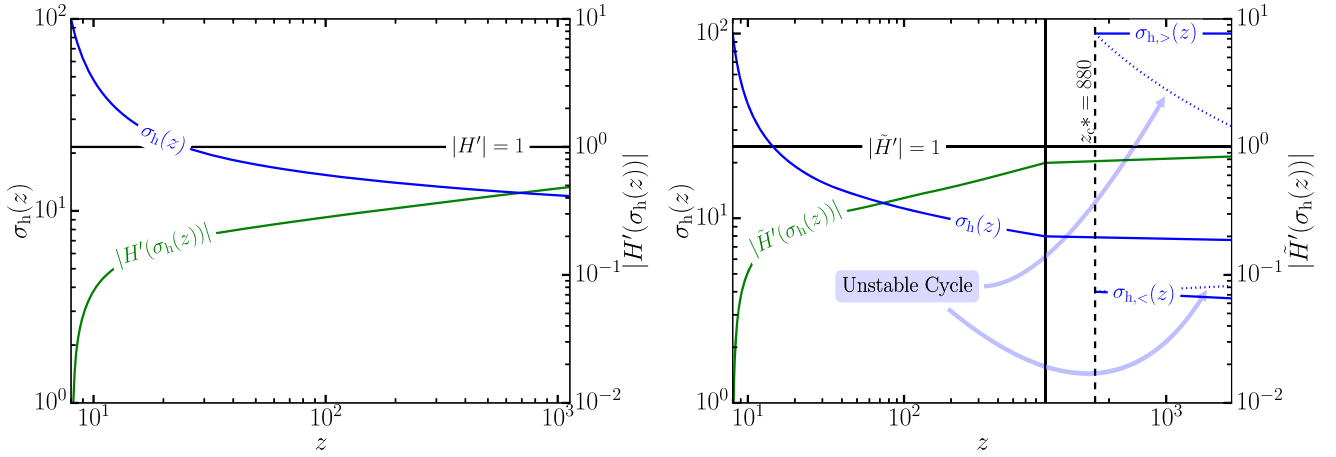

Figure B3. On the left (right), a stability analysis of the function $H(\sigma_h)$ [$\tilde{H}(\sigma_h)$] for all z between 8 and 1100. For visual clarity on the right plot, the horizontal axis scale beyond $z = 800$ is highly zoomed (and thus the kinks in curves crossing $z = 800$ are not real). The model without radiation back reaction on the layer particles does not possess high enough efficiency for its fixed point $\sigma_h(z)$ to go unstable, or to develop a two-state swing cycle. In contrast, the model with radiation reaction displays limit-cycle behaviour. There, pair feedback is so efficient that the system immediately starts, from the first readvection time, in a swing cycle whenever $z > z_{c*} = 880$. This is true even though the fixed point $\sigma_h(z)$ does not become locally unstable for any $z < 1100$. Instead, an *unstable* cycle appears and intercepts the flow on the iterated map, blocking all states that start from $\sigma_{h,0}$ from ever reaching the fixed point. Thus, the model on the right illustrates our argument in Appendix D that a locally unstable fixed point is a sufficient but not necessary condition for the system to globally asymptote to a limit cycle.

by the high-energy particles. These particles have their radiative cooling somewhat suppressed by Klein–Nishina effects and, for high enough z, may retain a large portion of the initially injected energy, substantially increasing f_{nocool}. Thus, as σ_h grows, one observes the solution $H(\sigma_h)$ [$\tilde{H}(\sigma_h)$] cross progressively larger constant-ξ contours of $h(\sigma_h)$ [$\tilde{h}(\sigma_h)$]. In fact, when radiation reaction on the layer particles is taken into account, $\tilde{\Xi}$ grows large enough at high z to initiate a dramatic two-state swing cycle – a very stark difference from when ξ is pessimistically estimated as $f_{\text{nocool, min}} \sim 0.01$.

In Fig. B2, we illustrate how the picture changes at smaller z. Here, the monotonic dependence of ξ on σ_h from Fig. B1 remains but is weaker. In particular, the largest value that ξ reaches, which occurs at high-σ_h, diminishes as one reduces z. Thus, a high cut-off in the layer particle energy distribution is needed to achieve order-unity efficiency. However, it is *not* the case that a lower cut-off causes ξ to plummet. As long as $z \geq 8$, and consistent with equation (B20), all models have appreciable efficiency, with $\xi \sim 0.1 \gg f_{\text{nocool,min}}$. This implies an important conclusion: even for a radiating layer particle

distribution that cuts off barely beyond $z = 8$, establishing a universal steady state is still possible for finite $\sigma_{h,0}$. The initial magnetization must only be greater than $\sim 1/\min(\xi) \sim 10$.

Finally, in Fig. B3, we present a complete stability analysis in z between 8 and 1100. This includes both how the fixed point solution $\sigma_h(z)$ varies with z, and whether a two-state swing cycle appears, with respective high and low magnetizations $\sigma_{h,>}(z)$ and $\sigma_{h,<}(z)$. The model that neglects radiation reaction on the layer particles never develops oscillatory behaviour, but always converges towards the fixed points $\sigma_h(z)$. (However, this could change if z is made even larger. In that case, layer particles with energies $> 4\tilde{\gamma}_{cool,2}$ spawn upstream pairs that retain nearly *all* of their energy while travelling back towards the layer, increasing ξ, and likely raising $|H'(\sigma_h)|$ above 1 in Fig. B3.) However, in the more radiatively self-consistent model, a swing cycle develops at a critical $z = z_c^* \simeq 880$. Intriguingly, this is before the fixed point $\sigma_h(z)$ becomes unstable, which does not happen for any $z < 1100$. Instead, an *unstable* two-cycle appears and intercepts the flow of the iterated map $\sigma_{h,n+1} = \tilde{H}(\sigma_{h,n})$, blocking states that start at $\sigma_{h,0}$ from reaching the fixed point.

While swing cycles appear to require quite extended particle distributions in our model ($z_c* = 880$ is close to the upper limit $z = 1100$ that we study), we do not think that this necessarily precludes their operation in all but the most extreme systems. Instead, we have found that the particular value of z where swing cycles set in is highly sensitive to order-unity changes in ξ, and, thus, this value could be either much higher (in which case swing cycles would be unlikely in reality) or much lower (in which case they may be quite common).

B4 Summary of self-consistent f_{nocool}

In this discussion, we have seen how, by self-consistently determining $\xi = f_{nocool}$ (through the functions Ξ and $\tilde{\Xi}$), one may achieve significantly higher efficiencies than a simple and pessimistic estimate, $\xi = f_{nocool, min} = 3/400$, would predict. Rather than radiating away most of the energy it receives from the reconnection layer, the newly created upstream plasma may catch and hold on to this energy, delivering an appreciable portion of it back to the layer. The reason is not that the fresh pairs radiate inefficiently. On the contrary, given a full readvection time, all of them would cool to energies $\sim \tilde{\gamma}_{cool,1}$. Rather, the broad distribution of newborn pairs is constantly being replenished, due to injection from photon annihilation, at high energies, and this allows the typical energy of radiating particles, as they enter the layer, to be quite high.

Thus, at some finite large (but not excessively large) initial magnetization $\sigma_{h,0} \ll 1/\min(\xi) \sim 10$, the layer *always* has a universal fixed point solution. Furthermore – and depending on how significantly radiation back reaction modifies the distribution of layer particles – energy recapture by the layer may reach order-unity efficiency when a long high-energy tail of layer particles extends beyond γ_{KN}. In this case, the reconnection layer may overshoot its fixed point steady state, and undergo late-time limit-cycle oscillations about this solution rather than converge towards it.

APPENDIX C: THE PARTICLE ESCAPE FACTOR f_{NOESC}

In this appendix, we discuss a complementary channel, besides radiative cooling, through which newly created upstream matter may lose energy: particle escape. Escape occurs when newborn pairs stream a distance $\sim L$ along the unreconnected magnetic field, vacating the system. Pairs can be born with essentially any pitch angle, so the typical escape time is $\sim L/c$ (except for a few particles

with very small pitch angles). Therefore, the escape factor f_{noesc} is close to unity if the readvection time, t_{ra}, satisfies $t_{ra} \simeq 10L/\tau_{\gamma\gamma}c < L/c$, which requires $\tau_{\gamma\gamma} > 10$.

Beyond this simple consideration, there are also other, more complicated kinetic effects that may influence the value of f_{noesc}. Many of these effects, at the same time, also help to decide whether above-threshold photons radiated from the layer ever reach the upstream region, as assumed throughout this work [e.g. assumption (iii)(b)]. These are, in principle, independent concepts: f_{noesc} pertains to pairs that have already been born into the upstream region; photon escape from the layer to the upstream region concerns the emission and propagation of radiation before pairs are ever produced. However, the same processes dictate both, and we discuss these parallel influences simultaneously.

Consider first the shape of the unreconnected magnetic field lines at a transverse distance $|y| < 0.1L$ from the layer. In this region, the unreconnected field is perturbed by the presence of large plasmoids, the largest of which may extend a distance $\sim 0.1L$ into the upstream region (Uzdensky et al. 2010; Sironi et al. 2016). Because the field is asymptotically uniform as $|y| \to \infty$, it is necessarily stronger (i.e. compressed), when $|y| \lesssim 0.1L$, in regions above plasmoids. This may lead to a magnetic bottling effect, where particles produced at $|y| < 0.1L$ tend to be mirror-confined between large plasmoids (even when they might otherwise escape), pushing f_{noesc} closer to unity. Unlike the effects discussed below, this confinement mechanism only pertains to particles already born and, therefore, does not affect what fraction of photons escape the layer to the upstream region.

We move now to a separate issue: anisotropy in the distribution of radiating layer particles. The layer tends to drive bulk plasma motion and, through the kinetic beaming mechanism (e.g. Cerutti et al. 2012b, 2013; Mehlhaff et al. 2020), collimated bunches of high-energy particles, into the $\pm x$-directions (either to the left or to the right in Figs 1 and 9). This biases the emitted photons (which are relativistically beamed along the directions of their emitters) into these same directions, and may, in turn, increase f_{noesc}. Suppose, for example, that a given above-threshold photon travels at an angle θ from the reconnection mid-plane. This reduces the total distance it propagates into the upstream region from $|y| \sim \lambda_{mfp}$ to $|y| \sim \lambda_{mfp}\sin\theta$. The readvection time for the produced pair is therefore reduced by the same factor: $\sin\theta$. Meanwhile, the escape time is still $\sim L/c$. Thus, the readvection time decreases relative to the escape time. This increases the fraction f_{noesc} of particles captured by the reconnection layer.

On the other hand, if the beaming of plasma motion (both on bulk and kinetic levels) is strong enough, the angle θ discussed above could be so small that most photons do not cross the separatrices into the upstream plasma before annihilating. This would occur if the transverse propagation distance $|y| \sim \lambda_{mfp}\sin\theta$ is smaller than the thickness Δ of the layer radiation zone (discussed in more detail below). In such a scenario, the created pairs would not load the inflow plasma and thus the pair-regulation mechanism would be somewhat suppressed.

We next discuss how the thickness, Δ, of radiation zones, where most pair-producing photons are emitted (discussed in Section 4 and Fig. 9), influences the escape of photons to the upstream region and f_{noesc}. If most photons are radiated from inside large plasmoids, instead of from thin strip-like radiation zones (of the kind argued for in Section 4), then the condition $\tau_{\gamma\gamma} > 10$, although it maintains $t_{ra} < L/c$ and helps keep f_{noesc} near unity, also confines most of the above-threshold photons to within the separatrices: inside plasmoids exceeding λ_{mfp} in size. This inhibits pair feedback. If, instead, the primary radiation sites are thin, even kinetic-scale, current sheets (e.g.

with $\Delta \ll \lambda_{\mathrm{mfp}}$, as we argue in Section 4) most photons travel to the upstream region and effectively load its plasma when they annihilate (modulo potential extreme beaming effects described above).[9]

Finally, let us consider how reconnection current sheets between merging plasmoids play into this picture. These miniature reconnection layers are oriented perpendicularly with respect to the main reconnection current sheet and therefore bias the motion of particles and emitted photons in the transverse $\pm y$-directions. The produced upstream pairs then have suppressed pitch angles (the $\pm y$ direction is perpendicular to the unreconnected magnetic field), which inhibits particle escape. However, if $\tau_{\gamma\gamma} > 10$, then not all radiation produced at merging-plasmoid current sheets escapes back into the inflow plasma: there will at least be some photon-confining plasmoids larger than $\lambda_{\mathrm{mfp}} < L/10$. Mergers between such plasmoids would not source significant self-regulating pair production. This, however, does not preclude merging-plasmoid reconnection from regulating itself (just on a smaller scale) in the same way that the main reconnection layer does as a whole.

In light of this discussion, there are clearly several details, influencing both f_{noesc} and the delivery of photons to the upstream region, that are beyond the scope of our present work to calculate quantitatively. These will require a future computational study in order to properly diagnose. Nevertheless, many of these effects (especially when $\tau_{\gamma\gamma} > 10$) promote order-unity f_{noesc}, and – modulo extreme beaming near reconnection X-points – allow photon escape to the upstream region from the primary large-scale current sheet.

APPENDIX D: GLOBAL STABILITY OF THE ITERATED MAP $x_{n+1} = h(x_n)$

In this section, the phrase 'late time' or 'late times' refers to the limit $\lim_{n \to \infty} t_n$ where $t_n \equiv n\lambda_{\mathrm{mfp}}/\beta_{\mathrm{rec}}c$. Late times defined in this sense are not necessarily reached by the reconnection system before reconnection terminates.

Here, we argue that the condition deciding the local stability of the iterated magnetization map $\sigma_{\mathrm{h},n+1} = h(\sigma_{\mathrm{h},n})$ from Section 4.1 also reveals, in some cases, its global stability. In particular, we argue that local instability $|h'(\sigma_{\mathrm{h}})| > 1$ implies that a system starting at magnetization $\sigma_{\mathrm{h},0}$ asymptotically approaches a two-state swing cycle. If, on the other hand, the fixed point is stable, $|h'(\sigma_{\mathrm{h}})| < 1$, then the system may or may not converge towards it. Thus, local instability is sufficient, but not necessary, for global instability.

We focus here only on strictly decreasing functions $h(x)$ [such that $h(y) < h(x)$ if $y > x$] that map into their own domain $[1, \sigma_{\mathrm{h},0}]$ (i.e. $h: [1, \sigma_{\mathrm{h},0}] \to [1, \sigma_{\mathrm{h},0}]$). Here, the upper end of the domain happens to coincide with the starting point of the map $\sigma_{\mathrm{h},0}$. The particular functions $h(x)$ and $H(x)$ specified in equations (51) and (B22) fall into this class. [Note here that we are overloading the symbol $h(x)$, designating with it a general class of functions and not necessarily the particular form in equation 51.] The arguments in this section can be generalized to non-increasing functions $h^*(x)$ such that $h^*(y) \leq h^*(x)$ if $y > x$ [into which category fall the functions $\tilde{h}(x)$ and $\tilde{H}(x)$ in equations 55 and B23] without difficulty, but require additional edge cases (saturation of the inequalities) to be considered, and so we do not formally treat them here.

We begin with a few basic observations. First, if the map does not start on the unique fixed point σ_{h}, then each successive iteration lands

on the opposite side of the fixed point from the preceding iteration. That is, if $\sigma_{\mathrm{h},n} < \sigma_{\mathrm{h}}$, then $\sigma_{\mathrm{h},n+1} > \sigma_{\mathrm{h}}$, and if $\sigma_{\mathrm{h},n} > \sigma_{\mathrm{h}}$, then $\sigma_{\mathrm{h},n+1} < \sigma_{\mathrm{h}}$. This is because

$$\sigma_{\mathrm{h},n+1} = h(\sigma_{\mathrm{h},n}) < h(\sigma_{\mathrm{h}}) = \sigma_{\mathrm{h}}, \tag{D1}$$

where the '<' follows when $\sigma_{\mathrm{h},n} > \sigma_{\mathrm{h}}$ by the strictly decreasing hypothesis. The same proof, but with a '>' sign, follows when $\sigma_{\mathrm{h},n} < \sigma_{\mathrm{h}}$. This precludes all non-fixed-point odd-period orbits. Hence, one cannot automatically infer the presence of chaos in the system using the three-period theorem (Li & Yorke 2004). As we now show, the dynamics are even more constrained: chaotic behaviour is, in fact, completely precluded.

For our second observation, we note that if, for any starting index m, one discovers that $\sigma_{\mathrm{h},m+2} < \sigma_{\mathrm{h},m}$, then for all integers $n \geq 1$, it is necessarily the case that $\sigma_{\mathrm{h},m+2n} < \sigma_{\mathrm{h},m+2(n-1)}$. That is, even though each successive iteration bounces to the opposite side of the fixed point, every consecutive even (or odd) iteration moves strictly in one direction. This follows inductively because

$$\sigma_{\mathrm{h},m+3} = h(\sigma_{\mathrm{h},m+2}) > h(\sigma_{\mathrm{h},m}) = \sigma_{\mathrm{h},m+1}$$
$$\Rightarrow \sigma_{\mathrm{h},m+4} = h(\sigma_{\mathrm{h},m+3}) < h(\sigma_{\mathrm{h},m+1}) = \sigma_{\mathrm{h},m+2}. \tag{D2}$$

The first line follows from the strictly decreasing hypothesis and the starting assumption that $\sigma_{\mathrm{h},m+2} < \sigma_{\mathrm{h},m}$. The second line follows from the strictly decreasing hypothesis and the first line.

This is a very powerful constraint because it severely limits the potential late-time dynamics of the system. In effect, the system can only either approach the fixed point $\sigma_{\mathrm{h}} = h(\sigma_{\mathrm{h}})$ or a two-state swing cycle. To see this, let us suppose that $\sigma_{\mathrm{h},2} < \sigma_{\mathrm{h},0}$. [Note that $\sigma_{\mathrm{h},2} = \sigma_{\mathrm{h},0}$ implies that the system begins in a two-cycle, and the condition $\sigma_{\mathrm{h},2} > \sigma_{\mathrm{h},0}$ is impossible because $h(x)$ maps on to its own domain.] Then each successive $\sigma_{\mathrm{h},2n}$ marches resolutely towards smaller values. This continuing reduction in $\sigma_{\mathrm{h},2n}$ can only be terminated in one way. It must be the case that there exists some $\sigma_{\mathrm{h},\infty}$ such that $\sigma_{\mathrm{h}} \leq \sigma_{\mathrm{h},\infty} = \lim_{n \to \infty} \sigma_{\mathrm{h},2n}$. If $\sigma_{\mathrm{h},\infty} = \sigma_{\mathrm{h}}$, then the system approaches the fixed point σ_{h} in a late-time steady state. If $\sigma_{\mathrm{h},\infty} > \sigma_{\mathrm{h}}$, then the system approaches a two-cycle. Note that it is not possible for the system to approach any $2n$-cycle for $n > 1$ that is not also a two-cycle. If that were the case, then one of the $\sigma_{\mathrm{h},n}$'s on the cycle would be on the same side of the fixed point as another $\sigma_{\mathrm{h},n}$ of the same parity (even or odd), and the two would not be equal. This would contradict our result that $\sigma_{\mathrm{h},2n+2} < \sigma_{\mathrm{h},2n}$.

Thus, there are only two possible asymptotic behaviours of the map $\sigma_{\mathrm{h},n+1} = h(\sigma_{\mathrm{h},n})$: either $\sigma_{\mathrm{h},n}$ approaches the fixed point σ_{h} or the map converges to a two-state cycle, and one of the states has magnetization $\sigma_{\mathrm{h},>} > \sigma_{\mathrm{h}}$. It follows that the flow of the iterated map from $\sigma_{\mathrm{h},0}$ to σ_{h} is always intercepted by a two-cycle when the local instability criterion $|h'(\sigma_{\mathrm{h}})| > 1$ is met. Otherwise, there would not be a suitable attractor (i.e. one consistent with both equations D1 and D2) to catch the strictly decreasing flow of the map $\sigma_{\mathrm{h},n+2} = h(h(\sigma_{\mathrm{h},n}))$. Let us illustrate these remarks with an example.

We display the particular map $\sigma_{\mathrm{h},n+2} = h(h(\sigma_{\mathrm{h},n}))$, where h is defined as in equation (51), in Fig. D1. For low ξ (in this case, lower than $\xi_{\mathrm{c}} = 0.84$), all of the $h(h(\sigma_{\mathrm{h},n}))$ curves intersect the diagonal line only once and with a gentle slope $|\mathrm{d}h(h(\sigma_{\mathrm{h},n}))/\mathrm{d}\sigma_{\mathrm{h},n}| < 1$. These intersection points coincide with the stable fixed points σ_{h} illustrated in Fig. 12. The fixed point becomes unstable when the slope of the twice-iterated map becomes tangent to the diagonal at a critical value $\xi_{\mathrm{c}} \simeq 0.84$. For ξ higher than this value, the map must intercept the diagonal in at least two other locations – the values $\sigma_{\mathrm{h},<}$ and $\sigma_{\mathrm{h},>}$ corresponding to a two-cycle. This is demanded in order for $h(h(x))$ to be at or below the diagonal as $\sigma_{\mathrm{h},n}$ approaches $\sigma_{\mathrm{h},0}$ and to be at or

---

[9]Note that, if $\lambda_{\mathrm{mfp}} > 0.1L$, it is irrelevant whether most of the pair-producing photons come from thin structures or from large round plasmoids. Either way, the photons escape to the upstream region.
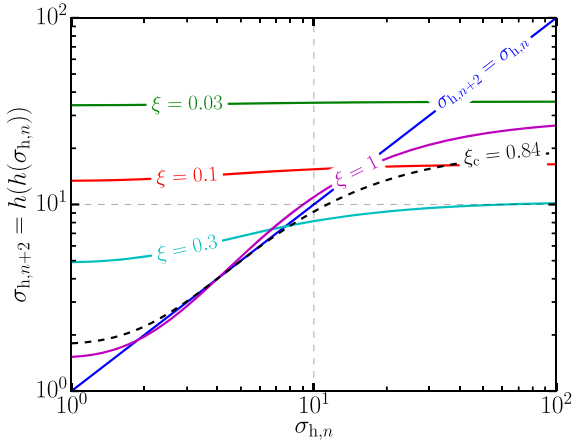
Figure D1. The twice-iterated map $\sigma_{h,n+2} = h(h(\sigma_{h,n}))$ with $h(x)$ as defined in equation (51). The appearance of a stable two-cycle coincides with the critical value $\xi_c \simeq 0.84$ where the fixed point at σ_h becomes unstable. This is consistent with the fact that there must be an attractor on the map $\sigma_{h,n+1} = h(\sigma_{h,n})$ to intercept the flow $\sigma_{h,2n+2} < \sigma_{h,2n} \leq \sigma_{h,0}$ when the local instability criterion $|h'(\sigma_h)| > 1$ is met. If locally unstable, the fixed point cannot be naked: it must be blocked from the initial state $\sigma_{h,0}$ by a two-state cycle.

above the diagonal as $\sigma_{h,n}$ approaches 1 (which itself is required by the fact that h maps into its own domain).

We stress that this general behaviour – where a naked fixed point σ_h is immediately concealed behind a two-cycle as soon as it goes unstable – is demanded by equations (D1) and (D2). However, the inverse – that all stable fixed points occur in isolation from two-cycles – does not hold in general. The topological argument of the preceding paragraph, which was based on the fact that $h(h(\sigma_{h,0})) \leq \sigma_{h,0}$ and $h(h(1)) \geq 1$, does not preclude the possibility that a stable fixed point could spontaneously become enshrouded by an even number of two-cycles with alternating stabilities. In this case, the function $h(h(\sigma_{h,n}))$ would simply intersect the diagonal an even number of times on either side of the fixed point. The first of these intersections, counted as one moves outward from the fixed point to the boundaries of the domain (1 and $\sigma_{h,0}$) would be necessarily unstable (since $|dh(h(\sigma_h))/d\sigma_h| > 1$), the next would be stable, the one after unstable, and so on until the outermost stable two-cycle. This would furthermore be consistent with equations (D1) and (D2), ensuring that a stable two-cycle attracts the late-time dynamics. Although such a state of affairs may seem at first rather unlikely, this is actually precisely what occurs for the map $\tilde{H}$ defined in equation (B23). As shown in Fig. B3, the fixed point on that map becomes enclosed inside two 2-cycles without *ever* becoming locally unstable.

All in all, if the fixed point σ_h is unstable ($|h'(\sigma_h)| > 1$) then the system necessarily asymptotes to a two-cycle. On the other hand, if σ_h is stable, then whether the system approaches it at late times depends on whether a two-cycle is present to intercept the flow on the iterated map. If a stable two-cycle exists, σ_h is never reached; otherwise, the system converges to σ_h. Fixed points and two-cycles are the only allowed late-time behaviours for monotonically decreasing maps.

APPENDIX E: EXACT SOLUTION TO THE THOMSON-LIMIT ENERGY-ADVECTION EQUATION

In this section, we solve equation (B9) exactly subject to the initial condition $N_{\gamma\gamma}^{(n)}(\gamma, 0) = 0$. We evolve the solution for one readvection time $t_{ra} = \lambda_{mfp}/\beta_{rec}c$, which simulates the continual injection of particles as a parcel of plasma moves from $|y| \sim \lambda_{mfp}$ to the layer. The solution follows identically for all generations, so we simplify our notation to

$$\frac{\partial}{\partial t} N(\gamma, t) + \frac{\partial}{\partial \gamma} \left[\dot{\gamma}_T N(\gamma, t)\right] = Q\delta(\gamma - \gamma_{KN}), \quad (E1)$$

and write $-\dot{\gamma}_T = \gamma/t_{cool,T}(\gamma) = c\gamma^2/L\gamma_{cool} \equiv \gamma^2/\tau$, abbreviating $\tau \equiv L\gamma_{cool}/c$. (Note that τ is a time and not an optical depth.)

We denote the Laplace transform in time with a tilde:

$$\tilde{N}(\gamma, s) \equiv \int_0^\infty N(\gamma, t)e^{-st}dt. \quad (E2)$$

Laplace-transforming equation (E1) gives

$$s\tilde{N} - \frac{1}{\tau}\frac{\partial}{\partial \gamma}\left(\gamma^2\tilde{N}\right) = \frac{Q}{s}\delta(\gamma - \gamma_{KN}), \quad (E3)$$

where we used the initial condition $N(\gamma, 0) = 0$. Expanding the γ-derivative, rearranging, and multiplying through by the integrating factor

$$\mu(\gamma) = \gamma^2 e^{-s\tau(1-1/\gamma)} \quad (E4)$$

gives

$$\frac{\partial}{\partial \gamma}\left[\mu(\gamma)\tilde{N}(\gamma, s)\right] = -\frac{Q\tau}{s\gamma^2}\mu(\gamma)\delta(\gamma - \gamma_{KN}). \quad (E5)$$

Next, we integrate (E5) from γ to some arbitrary $\gamma_{hi} > \gamma_{KN}, \gamma$. Using $\tilde{N}(\gamma_{hi}, s) = 0$ (radiative cooling only populates energies lower than γ_{KN}) yields

$$\begin{aligned}\tilde{N}(\gamma, s) &= \frac{Q\tau}{s\gamma_{KN}^2}\frac{\mu(\gamma_{KN})}{\mu(\gamma)}\Theta(\gamma_{KN} - \gamma) \\ &= \frac{Q\tau}{s\gamma_{KN}^2}e^{-s\tau(1/\gamma - 1/\gamma_{KN})}\Theta(\gamma_{KN} - \gamma), \quad (E6)\end{aligned}$$

where $\Theta(x)$ is the Heaviside step function. Equation (E6) is the Laplace transform of our sought solution:

$$N(\gamma, t) = \frac{Q\tau}{\gamma^2}\Theta\left[t - \tau\left(\frac{1}{\gamma} - \frac{1}{\gamma_{KN}}\right)\right]\Theta(\gamma_{KN} - \gamma). \quad (E7)$$

Thus, constant injection at $\gamma = \gamma_{KN}$ develops into a power-law γ^{-2} extending into lower and lower energies with time. Putting $t = \lambda_{mfp}/\beta_{rec}c \simeq 10L/\tau_{\gamma\gamma}c$ allows the power law to extend from γ_{KN} all the way down to

$$\left(\frac{1}{\gamma_{KN}} + \frac{1}{\tilde{\gamma}_{cool,1}}\right)^{-1} \simeq \tilde{\gamma}_{cool,1} \equiv \frac{3}{50}\gamma_{KN}. \quad (E8)$$

Hence (cf. equation B10),

$$N(\gamma, t_{ra}) = Q\gamma_{cool}\frac{L}{c}\begin{cases}\gamma^{-2} & \tilde{\gamma}_{cool,1} \leq \gamma < \gamma_{KN} \\ 0 & \text{otherwise}\end{cases}. \quad (E9)$$

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