Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **517,** 5413–5448 (2022) https://doi.org/10.1093/mnras/stac2909
Advance Access publication 2022 October 12

Standard self-confinement and extrinsic turbulence models for cosmic ray transport are fundamentally incompatible with observations

Philip F. Hopkins [1]★ Jonathan Squire [2] Iryna S. Butsky[1] and Suoqing Ji[1]

[1]*TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[2]*Physics Department, University of Otago, 730 Cumberland St., Dunedin 9016, New Zealand*

Accepted 2022 October 5. Received 2022 October 5; in original form 2021 December 1

ABSTRACT

Models for cosmic ray (CR) dynamics fundamentally depend on the rate of CR scattering from magnetic fluctuations. In the ISM, for CRs with energies ∼MeV-TeV, these fluctuations are usually attributed either to 'extrinsic turbulence' (ET) – a cascade from larger scales – or 'self-confinement' (SC) – self-generated fluctuations from CR streaming. Using simple analytic arguments and detailed 'live' numerical CR transport calculations in galaxy simulations, we show that both of these, in standard form, cannot explain even basic qualitative features of observed CR spectra. For ET, any spectrum that obeys critical balance or features realistic anisotropy, or any spectrum that accounts for finite damping below the dissipation scale, predicts qualitatively incorrect spectral shapes and scalings of B/C and other species. Even if somehow one ignored both anisotropy and damping, observationally required scattering rates disagree with ET predictions by orders of magnitude. For SC, the dependence of driving on CR energy density means that it is nearly impossible to recover observed CR spectral shapes and scalings, and again there is an orders-of-magnitude normalization problem. But more severely, SC solutions with super-Alfvénic streaming are unstable. In live simulations, they revert to either arbitrarily rapid CR escape with zero secondary production, or to bottleneck solutions with far-too-strong CR confinement and secondary production. Resolving these fundamental issues without discarding basic plasma processes requires invoking different drivers for scattering fluctuations. These must act on a broad range of scales with a power spectrum obeying several specific (but plausible) constraints.

Key words: MHD – plasmas – turbulence – cosmic rays – ISM: structure – galaxies: evolution.

1 INTRODUCTION

Understanding how cosmic rays (CRs) propagate and interact as they travel through the inter-stellar medium (ISM) and circum/inter-galactic medium (CGM/IGM) is a problem with crucial implications for a wide variety of questions in astrophysics, including star, planet, and galaxy formation and evolution; astro-chemistry and chemo-dynamics; and space plasma physics (for reviews, see Zweibel 2013, 2017; Amato & Blasi 2018; Kachelrieß & Semikoz 2019). Most of the energy density in CRs (which determines their ability to ionize, heat, and interact with the gas) resides around ∼GeV energies, and in the range ∼MeV–TeV. At these energies, CR gyroradii ($r_{g,\rm cr} \sim 10^{9} - 10^{15}$ cm) are vastly smaller than the characteristic scale-lengths of the galactic disc and the driving scales of ISM turbulence. As such, CRs cannot simply 'free stream' out of the galaxy at speeds $\sim c$, but rather scatter in pitch angle from magnetic-field fluctuations, giving rise to some effective scattering rate $\nu_{\rm s}$. This, in turn, leads to bulk CR transport, which can be parametrized by some effective diffusivity $\kappa_{\rm eff} \sim c^2/\nu_{\rm s}$ or streaming speed $v_{\rm st} \sim \kappa_{\rm eff}|\nabla P_{\rm cr}|/P_{\rm cr} \ll c$.

These scattering rates have major implications for all of the astrophysical and space plasma fields above, and are probably the most important uncertainty in our understanding of the role of CRs. In understanding star and galaxy formation and the effect of CRs on the ISM/CGM/IGM, for example, a multitude of studies have shown that if the effective diffusivity or streaming speed is 'too small,' CRs will be trapped in dense gas, and rapidly lose their energy to a variety of processes (e.g. pionic, catastrophic, and synchrotron/inverse Compton losses) before they can have a significant effect on the gas properties. In the opposite limit, if the diffusivity is 'too large,' CRs will free-stream rapidly out of the CGM and either effectively decouple from the gas or build up so little pressure that they will again have no effect. But, at diffusivities 'in between' these values, ∼GeV CRs can have energy densities that are comparable to magnetic or thermal energy densities and have important effects on the gas (Girichidis et al. (Girichidis et al. 2016; Wiener, Pfrommer & Oh 2017; Butsky & Quinn 2018; Farber et al. 2018; Butsky et al. 2020; Su et al. 2020; Hopkins et al. 2020b, 2021c). Likewise, the effect of CRs on astro-chemistry, planet formation, and dense gas systems depends sensitively on how effectively very low-energy ($\lesssim$ 100 MeV) CRs are trapped and their penetration depth into dense clouds and protostellar discs (Wolfire et al. 1995; Scalo & Elmegreen 2004; Indriolo, Fields & McCall 2009; Padovani, Galli & Glassgold 2009; Thompson 2013; Lee & Hopkins 2020; Parker 2020; Bustard & Zweibel 2021). And details of CR plasma physics, in particular their micro–scale interactions with the multiphase ISM/CGM/IGM gas, both shape and are determined

★ E-mail: phopkins@caltech.edu

by the CR scattering rates (see e.g. Zank 2014; Bai et al. 2015, 2019; Lazarian 2016; Zweibel 2017; Holcomb & Spitkovsky 2019; Thomas & Pfrommer 2019; Van Marle, Casse & Marcowith 2019, and references therein)

The overwhelming majority of work studying and attempting to constrain CR transport, either in the Milky Way (MW) galaxy from observations in and around the Solar system (from e.g. terrestrial and space-based CR experiments) or from γ-ray observations, has focused on simple, phenomenological constraints. This generally involves parametrizing CR transport in terms of effective diffusion coefficients or streaming speeds or other transport parameters (see Strong & Moskalenko 2001; e.g. Blasi & Amato 2012; Vladimirov et al. 2012; Gaggero et al. 2015; Cummings et al. 2016; Guo, Tian & Jin 2016; Jóhannesson et al. 2016; Korsmeier & Cuoco 2016; Evoli et al. 2017, and references therein). These 'effective' coefficients represent, by definition, some weighted average in the ISM between CR sources (e.g. SNe remnants, in the MW) and the Solar system, and are often parametrized as e.g. a power-law function of the CR rigidity $R_{\rm cr}$, such as $\kappa_{\rm eff} = \kappa_0\,\beta_{\rm cr}\,(R_{\rm cr}/R_{\rm cr,\,0})^{\delta_s}$.

But Solar system constraints only measure CR transport in an average sense at one point in space and time, while ISM properties – both along the CR 'path' and in different galaxies and cosmic epochs – vary tremendously in both space and time (by many orders of magnitude for quantities of interest like magnetic energy density). Further, phenomenological models do not explain how such coefficients arise in the first place. What is therefore required is a *physical* model of CR transport that can reproduce these effective constraints and be tested in other regimes. However, this is particularly challenging at the MeV-TeV CR energies of greatest interest, because (1) the observational constraints are limited, (2) the extremely small gyro radii are much smaller than spatially resolvable scales in most astrophysical ISM studies, (3) the 'back reaction' of magnetic fields and gas from CRs, e.g. via gyroresonant instabilities and macroscopic CR 'pressure' effects, is maximized around this energy scale, and (4) the ISM, CGM, and IGM phase structure and turbulence itself remains uncertain.

Broadly speaking, historical models that attempt to *predict* CR scattering rates and transport parameters at these energies fall into one of two broad categories: 'extrinsic turbulence' (ET) and 'self-confinement' (SC) models. In the simplest ET models, going back to e.g. Jokipii (1966), Wentzel (1968), Skilling (1975a), and Voelk (1975), CRs scatter from gyroresonant fluctuations in $\mathbf{B}$, i.e. those with wavenumbers $k_\parallel \equiv \mathbf{k}\cdot\hat{\mathbf{b}} \sim 1/r_{g,\rm cr}$. Those early models assumed $\delta\mathbf{B}(k_\parallel)$ was sourced by an isotropic, undamped, inertial-range Kolmogorov (1941)-type (K41) cascade from larger ISM scales. This gives rise to a scattering rate $\nu_s \sim \Omega_{\rm cr}\,|\delta\mathbf{B}(k_\parallel)|^2/|\mathbf{B}|^2 \propto |\mathbf{B}|^{1/3}\,\ell_A^{2/3}\,R_{\rm cr}^{-1/3}$ where ℓ_A is the Alfvén scale of the cascade.[1] In SC models, going back to Wentzel (1969), Skilling (1971), and Holman, Ionson & Scott (1979), CRs themselves source the scattering modes, which they excite via various instabilities as they stream down magnetic field lines (Wentzel 1968; Kulsrud & Pearce 1969). The instabilities grow until reaching some saturation amplitude that is determined by a wave damping rate Γ, thus giving rise to scattering rates that scale as $\nu_s \sim \Omega_{\rm cr}\,v_A\,|\nabla P_{\rm cr}|/(\Gamma\,|\mathbf{B}|^2)$ (Skilling 1975b).

Until recently, it has not been possible to directly test and compare these models with local CR observations for a variety of reasons. Perhaps most importantly, even in the simplest ET and SC models, scattering rates are not constant but depend strongly on ISM properties. These, in turn, vary dramatically across the ISM by as much as $\sim$10 orders of magnitude, in a manner that cannot be captured by simplified models discussed above that assume some steady-state CR distribution and solve e.g. a 'leaky-box' or 'flat-halo diffusion' model with a simple analytic galaxy model (see the review in Hopkins et al. 2021d). Moreover, only recently has the fluid theory of CRs been developed to the point where SC theories can be 'coarse-grained' self consistently into fluid-like magnetohydrodynamics (MHD)-CR transport and scattering models (Zweibel 2013, 2017; Thomas & Pfrommer 2019; Hopkins, Squire & Butsky 2022b), while modern versions of SC and ET models that account for important effects such as damping and anisotropy have only been developed in the last two decades (Chandran 2000; Yan & Lazarian 2002; Farmer & Goldreich 2004; Yan & Lazarian 2004, 2008; Zweibel 2017; Squire et al. 2021). Finally, only recently has CR data become available from outside of the heliopause, which is crucial for the CRs of greatest interest ($\lesssim$ 100 GeV energies) because these are strongly modulated by the Sun (Cummings et al. 2016; Bindi et al. 2017; Bisschoff, Potgieter & Aslam 2019). These new observations help to remove the order-of-magnitude degeneracies that plagued previous attempts to test CR transport/scattering theories.

In this paper, we therefore revisit the question of whether or not state-of-the-art ET or SC models can possibly explain the state-of-the-art CR observations. We first consider the problem in a purely analytic fashion, synthesizing CR transport theories (beginning from general considerations before considering approximations such as steady-state behaviour) and reviewing the state of the art in both SC and ET theories in order to treat all potentially important damping terms. We then test these models in even greater detail with fully non-equilibrium, non-linear, non-steady state CR transport in high-resolution galaxy simulations, which explicitly resolve the plasma properties that determine CR scattering. While a first attempt at such comparisons was presented in Hopkins et al. (2021d), which already argued that present ET and SC models failed to reproduce the observations, that paper simplified by considering a 'single-bin' CR approximation, essentially modelling only CR protons in a narrow range of energies at $\sim$ 1 GeV. Here, we expand this to a full spectrum of CRs with a wide range of secondary species. This dramatically expands the range of observational constraints and will allow us to show that the scope of the discrepancy between SC and ET models and observations is much larger than previously believed. In particular, some of the possible resolutions to the discrepancies noted in Hopkins et al. (2021d) – e.g. changing the normalization of SC-induced scattering rates by accounting for certain pitch-angle effects – cannot possibly provide the full solution. We use these constraints to propose that a new class of sources for gyroresonant scattering waves is required, which obeys a well-constrained (but plausible) set of requirements.

In Section 2, we set up the analytic background, including review of some key definitions (Section 2.1) and description of the CR dynamics equations (Section 2.2), relevant Alfvén wave properties (Section 2.3), and expressions for scattering rates (Section 2.4). We then review standard damping mechanisms (Section 2.5) and drivers of scattering fluctuations in both SC (Section 2.6) and ET (Section 2.7) limits, and the resulting steady-state behaviours (Section 2.8). In Section 3, we discuss the problems that follow: first we review what empirical CR transport models require (Section 3.1) then describe how both ET (Section 3.2) and SC (Section 3.4) models cannot satisfy these constraints, then propose phenomeno-

[1] We define the 'Alfvén scale' ℓ_A of any large-scale turbulent cascade as the scale where extrapolating the inertial range, $\langle|\delta\mathbf{v}_{\rm turb}(k\sim 1/\ell_A)|\rangle \approx v_{A,\,\rm ideal}$ (the ideal Alfvén speed).

logical solutions (Section 3.5) involving either modified damping (Section 3.5.1) or driving (Section 3.5.2) terms. We then proceed to explore these in detailed simulations. Section 4 describes the numerical methods, outlining the non-CR (Section 4.1) and CR (Section 4.2) physics simulated, a theoretically motivated 'reference' model (Section 4.3), and extensive variations to that model that we have considered (Section 4.4). Section 5 describes the results of these simulations, first (Section 5.1) confirming the analytically predicted 'failure modes' of SC (Section 5.1.1) and ET (Section 5.1.2) models, then testing the proposed alternative damping (Section 5.2) or driving (Section 5.3) scalings to see if these can reproduce observations. We summarize and conclude in Section 6. Appendices A–C contain more detailed analytic derivations of steady-state CR behaviours and turbulent scalings.

2 ANALYTIC BACKGROUND

2.1 Key scales and definitions

To begin, we review some important concepts. Table 1 collects definitions of some of the most-commonly used variables in this paper. Per Section 1, CRs with some rigidity $R_{\rm cr}$ and corresponding gyro radius $r_{g,\rm cr} = R_{\rm cr}/|\mathbf{B}|$ ($\sim 10^{-6}\,$pc in the diffuse ISM, for CRs with $R_{\rm cr}\sim 1\,$GV) are scattered in pitch angle μ by fluctuations in the magnetic field $\delta\mathbf{B}$ with some effective scattering rate ν_s. In most models (though not all, as we discuss below), the CR scattering rate is strongly dominated by gyroresonant scattering of CRs from Alfvén waves with parallel wavenumbers $k_\parallel \equiv \mathbf{k}\cdot\hat{\mathbf{b}} \sim 1/r_{g,\rm cr}$. The power in these modes ($e_A \sim \langle|\delta\mathbf{B}(k_\parallel\sim 1/r_{g,\rm cr})|^2\rangle/8\pi$), which determines ν_s, is set by competition between some source/driving terms S and damping or dissipation rates $Q = \Gamma\, e_A$.

We stress that this encompasses both SC and ET models: the difference comes down to which dominates S. In SC models, S is sourced by parallel Alfvén waves excited directly by CRs (via e.g. gyroresonant and streaming instabilities), which we denote $S_{\rm sc}$. In ET models, the dominant contribution to S comes from a turbulent cascade $S_{\rm et}$ operating over a large dynamic range in scale. Also note that by definition Γ includes any terms which remove power from the scattering modes, e.g. both traditional collisional damping, but also processes which transfer energy to other modes with different wavenumbers or weaker scattering effects.

We will show that it is useful to parametrize S and Γ in terms of their approximate scaling with parallel wavenumber ($k_\parallel$), total kinetic energy density of CRs around a given rigidity ($\epsilon_{\rm cr}$), and energy in scattering modes at some $k_\parallel$ (e_A or $e_\pm$), as $S\propto k_\parallel^{\zeta_k}\, e_A^{\zeta_A}\,\epsilon_{\rm cr}^{\zeta_{\rm cr}}$ and $\Gamma\propto k_\parallel^{\xi_k}\, e_A^{\xi_A}\,\epsilon_{\rm cr}^{\xi_{\rm cr}}$ shown in Table 2.[2] The key qualitative problems and failure modes of SC and ET theories can be encapsulated entirely in these coefficients ($\zeta_k,\zeta_A,\zeta_{\rm cr},\xi_k,\xi_A,\xi_{\rm cr}$). Essentially, we will show that whether or not a theory of CR scattering can potentially reproduce CR observations (independent of normalization parameters) depends on these few numbers.

It is also helpful to recall some key scales in turbulence. Most of the power in ISM/CGM turbulence is on the driving scale, typically $\gtrsim 0.1-1\,$kpc (on which scale the turbulence is often trans or super-Alfvénic, $\langle|\delta\mathbf{v}_{\rm turb}(k)|^2\rangle^{1/2}\gtrsim v_{A,\rm ideal}$). Below the Alfvén scale ℓ_A (typically $\sim 10-100\,$pc in the ISM), the turbulent fluctuations are sub-Alfvénic ($\langle|\delta\mathbf{v}_{\rm turb}(k\gtrsim 1/\ell_A)|^2\rangle^{1/2}\lesssim v_{A,\rm ideal}$). At the vastly smaller gyro scale $r_{g,\rm cr}\ll\ell_A$, the scattering fluctuations are fractionally small (quasi-linear; $|\delta\mathbf{B}(k\sim 1/r_{g,\rm cr})|/|\mathbf{B}|\ll 1$), and we can treat fluctuations (approximately) as a superposition of Alfvén, slow, and fast magnetosonic modes.

Alfvén modes are only weakly damped down to scales much smaller than CR gyroresonant scales (at least down to ion gyroradii). When we refer below to 'damping' terms acting directly on the CR scattering modes ($Q_\pm$ and $\Gamma_\pm$), we generally are referring to this 'weak' damping. Specifically, Alfvén-mode-damping times ($\sim\Gamma^{-1}$) at some k are much longer than the mode-crossing times $\sim 1/(k\,v_{A,\rm eff})$ (by typical factors $\sim 10^4$–10^8). However, as we discuss below (and in more detail in Appendix C), it is well-established that an Alfvénic (or slow magnetosonic) cascade must be highly anisotropic on scales below the Alfvén scale ℓ_A ($k\gg 1/\ell_A$): an isotropic Iroshnikov (1963), Kraichnan (1965)-type (IK) cascade, for example, simply cannot exist (it is not mathematically self-consistent) on scales $r_{g,\rm cr}\ll\ell_A$. This means it is crucial to distinguish between parallel $k_\parallel$ and perpendicular components of $\mathbf{k}$.

Fast magnetosonic modes, on the other hand, are orders of magnitude more strongly damped on small scales by both collisional and collisionless/Landau damping (see Appendix C). On scales below the dissipation scale $k_{\rm diss}\sim 1/\ell_{\rm diss}$ (with typical $\ell_{\rm diss}\gtrsim 0.001\,$pc in the ISM), the magnetosonic mode damping time $\Gamma^{-1}_{\rm magnetosonic}$ becomes shorter than turbulent 'cascade' or decoherence or energy-transfer time-scale $\tau_{\rm cas}$ (with e.g. $\tau_{\rm cas}^{-1}(k)\sim k\,\langle|\delta\mathbf{v}^2_{\rm turb}(k)|\rangle^{1/2}$ in the classical K41 picture), so the cascade must be truncated or strongly modified by the energy losses. For essentially all plausible ISM/CGM conditions, the gyroresonant scales are much smaller than the dissipation scale ($r_{g,\rm cr}\ll\ell_{\rm diss}$) at rigidities $\lesssim 100-1000\,$GV, so one cannot simply extrapolate an un-damped inertial-range magnetosonic cascade of any form (let alone K41) down to gyroresonant scales.

One additional clarification is important. To be consistent with the previous literature, when we refer to the 'turbulent damping' of gyroresonant modes, $\Gamma_{\rm turb}$ (see Section 2.5), we refer specifically to a process by which interactions between gyroresonant scattering modes and other turbulent modes transfer energy from the weakly damped gyroresonant scattering modes (with $k_\parallel\sim 1/r_{g,\rm cr}$) to either higher-$k$ or more strongly damped modes. This is different from 'damping or dissipation of turbulence,' which we will use to refer to the phenomena described above for e.g. fast magnetosonic modes, in which a turbulent cascade is strongly modified by sufficiently strong damping on some dissipation scale $\ell_{\rm diss}$ larger than the gyroresonant scales.

2.2 Cosmic ray dynamics equations

Consider an arbitrary CR distribution function (DF) $f_{\rm cr} = f_{\rm cr}(\mathbf{x},\,\mathbf{p}_{\rm cr},\,t,\,s_{\rm cr},\,\ldots)$ as a function of position $\mathbf{x}$, CR momentum $\mathbf{p}_{\rm cr}$, time t, and CR species $s_{\rm cr}$, on macroscopic scales much larger than CR gyroradii. Assuming the DF is approximately gyrotropic, and the background gas velocities $\mathbf{u}_{\rm gas}$ are non-relativistic, the general Vlasov equation for $f_{\rm cr}$ can be written to leading order in $\mathcal{O}(|\mathbf{u}_{\rm gas}|/c)$ as the usual focused transport equation (Skilling 1971, 1975a; Isenberg 1997; Le Roux, Matthaeus & Zank 2001; Le Roux et al. 2005; Zank 2014; Le Roux et al. 2015), with the standard quasi-linear theory slab scalings for the scattering terms from Schlickeiser (1989). As shown in Hopkins et al. (2022b), taking the zeroth and first pitch-angle (μ) moments of that equation (retaining all terms to leading order in

[2]Note that $k_\parallel$, $\epsilon_{\rm cr}$, and e_A do not need to be strictly independent variables for this parametrization.

Table 1. Commonly used variables in this paper.

$f_{\rm cr}$, μ, $p_{\rm cr}$, $E_{\rm cr}$	CR DF $f_{\rm cr} \equiv f_{\rm cr}(\mathbf{x},\ \mathbf{p}_{\rm cr},\ t,\ s,\ \ldots)$, pitch angle $\mu \equiv \hat{\mathbf{p}}_{\rm cr}\cdot\hat{\mathbf{b}}$, momentum $p_{\rm cr} = \lvert\mathbf{p}_{\rm cr}\rvert$, energy $E_{\rm cr}$
$\Omega_{\rm cr}$, $v_{\rm cr}$, $r_{g,\,\rm cr}$, $R_{\rm cr}$	CR gyrofrequency $\Omega_{\rm cr}$, velocity $v_{\rm cr} = \beta_{\rm cr}\,c$, gyroradius $r_{g,\,\rm cr} \equiv v_{\rm cr}/\Omega_{\rm cr}$, rigidity $R_{\rm cr}$
$j_{\rm cr}(R_{\rm cr})$	CR injection rate/spectrum as a function of rigidity $R_{\rm cr}$ (from SNe and other sources)
$e_{\pm}$, e_A	Energy of forward(+) or backward(-) CR-scattering waves $e_{\pm} \equiv k_{\parallel}\mathcal{E}(\mathbf{k}\cdot\hat{\mathbf{b}} = \pm k_{\parallel})$ at wavenumber $\pm k_{\parallel}$, with $e_A \equiv e_+ + e_-$
$\nu_{\rm s,\,\pm}$, $\bar{\nu}_{\rm s,\,\pm}$	CR scattering rate from forward/backward propagating waves $\nu_{\rm s,\,\pm}$, and pitch-angle averaged $\bar{\nu}_{\rm s,\,\pm}$
$\delta_{\rm s}$	Average dependence of CR scattering rate on rigidity, e.g. $\bar{\nu}_{\rm s} \propto \beta_{\rm cr}\,R_{\rm cr}^{-\delta_{\rm s}}$ (observationally required $0.4 \lesssim \delta_{\rm s} \lesssim 0.7$)
$\mathbf{k}$, $k_{\parallel}$, $k_{\perp}$	Wavenumber $\mathbf{k}$ of CR-scattering modes, with parallel ($k_{\parallel} \equiv \mathbf{k}\cdot\hat{\mathbf{b}}$) and perpendicular ($k_{\perp}$) components
$e'_{\rm cr}$, $\epsilon_{\rm cr}$, $P'_{\rm cr}$	Differential CR energy density/pressure at a given momentum $p_{\rm cr}$, $e'_{\rm cr} \equiv de_{\rm cr}/d\ln p_{\rm cr}$, $\epsilon_{\rm cr} = (\gamma_{\rm cr}-1)\,e'_{\rm cr}$, $P'_{\rm cr} = \beta_{\rm cr}^2\,e'_{\rm cr}/3$
$\mathbf{B}$, $\hat{\mathbf{b}}$, $e_{\rm B}$, $\delta\mathbf{B}$	Magnetic field $\mathbf{B}$, direction $\hat{\mathbf{b}} \equiv \mathbf{B}/\lvert\mathbf{B}\rvert$, energy $e_{\rm B} \equiv \lvert\mathbf{B}\rvert^2/8\pi$, fluctuations $\delta\mathbf{B}$ on scale $\sim k$
$v_{A,\,\rm ideal}$, $v_{A,\,\rm eff}$	Ideal-MHD Alfvén speed $v_{A,\,\rm ideal} \equiv (\lvert\mathbf{B}\rvert^2/4\pi\rho)^{1/2}$, speed of gyroresonant Alfvén waves $v_{A,\,\rm eff}$ (equation 5)
ℓ_A	Alfvén scale of large-scale turbulence (scale where $\langle\lvert\delta\mathbf{v}_{\rm turb}(k\sim 1/\ell_A)\rvert^2\rangle^{1/2} \sim v_{A,\,\rm ideal}$)
$S_{\pm}$	Source terms for CR scattering modes ($D_t e_{\pm} + \ldots = S_{\pm}$)
$Q_{\pm}$, $\Gamma_{\pm}$	Damping terms for CR scattering modes ($D_t e_{\pm} + \ldots = -Q_{\pm} \equiv -\Gamma_{\pm}\,e_{\pm}$)

Table 2. Parametrization of generalized damping/driving rates for CR-scattering modes.

ξ_k, ξ_A, $\xi_{\rm cr}$	Coefficients for damping rates: $Q_{\pm} \equiv \Gamma_{\pm}\,e_{\pm}$ with $\Gamma_{\pm} \propto k_{\parallel}^{\xi_k}\,e_{\pm}^{\xi_A}\,\epsilon_{\rm cr}^{\xi_{\rm cr}}$	ξ_k	ξ_A	$\xi_{\rm cr}$
	Values explored in our simulation survey	$0 \le \xi_k \le 2$	$0 \le \xi_A \le 1$	$0 \le \xi_{\rm cr} \le 1$
$X_{\rm in}$	Quantities (e.g. Q, Γ, ξ) for ion-neutral damping	0	0	0
$X_{\rm dust}$	Quantities for dust damping	$0.5 \rightarrow 0.75$	0	0
$X_{\rm nll}$	Quantities for NLL damping	1	1	0
$X_{\rm turb/LL}$	Quantities for linear Landau or 'turbulent' damping	$0.4 \rightarrow 0.5$	0	0
$X_{\rm new,\,damp}$	Quantities for proposed novel damping that could fit observations	$\sim 0.1 \rightarrow 0.4$	~ 0	~ 1
ζ_k, ζ_A, $\zeta_{\rm cr}$	Coefficients for driving/source rates: $S_{\pm} \propto k_{\parallel}^{\zeta_k}\,e_{\pm}^{\zeta_A}\,\epsilon_{\rm cr}^{\zeta_{\rm cr}}$	ζ_k	ζ_A	$\zeta_{\rm cr}$
	Values explored in our simulation survey	$-2 \lesssim \zeta_k \lesssim 2$	$0 \le \zeta_A \le 1$	$0 \le \zeta_{\rm cr} \le 1$
$X_{\rm sc}$	Quantities (e.g. S, ζ) for SC driving (non-steady-state ζ in [])	0 [1]	0 [1]	1 [1]
$X_{\rm et}$	Quantities for ET driving (with anisotropy/damping)	$\lesssim -1 + \xi_k^{\rm turb}$	~ 0	0
$X_{\rm new,\,lin}$	Quantities for proposed novel linear source terms which could fit observations	$\sim 0.6 \rightarrow 0.9$	~ 1	~ 0
$X_{\rm new,\,ext}$	Quantities for proposed novel extrinsic source terms which could fit observations	$\sim -0.25 \rightarrow -0.1$	~ 0	~ 0

$\mathcal{O}(\lvert\mathbf{u}_{\rm gas}\rvert/c)$) gives the evolution equations for the isotropic part of the DF $\bar{f}_{\rm cr,\,0}$ (i.e. the CR number density at a given differential $p_{\rm cr} = \lvert\mathbf{p}_{\rm cr}\rvert$) and its flux:[3]

$$D_t\,\bar{f}_{\rm cr,\,0} + \nabla\cdot\left(v_{\rm cr}\,\hat{\mathbf{b}}\,\bar{f}_{\rm cr,\,1}\right) = j_{\rm cr,\,0} + \frac{1}{p_{\rm cr}^2}\frac{\partial}{\partial p_{\rm cr}}\left[p_{\rm cr}^2\left\{\mathcal{R}_{\rm loss}\,\bar{f}_{\rm cr,\,0} + \left(\mathbb{D}_{\rm cr}:\nabla\mathbf{u}_{\rm gas}\right)\,p_{\rm cr}\,\bar{f}_{\rm cr,\,0} + \tilde{D}_{p\mu}\,\bar{f}_{\rm cr,\,1} + \tilde{D}_{pp}\,\frac{\partial\bar{f}_{\rm cr,\,0}}{\partial p_{\rm cr}}\right\}\right]$$

$$D_t\,\bar{f}_{\rm cr,\,1} + \hat{\mathbf{b}}\cdot\left[\nabla\cdot\left(v_{\rm cr}\,\mathbb{D}_{\rm cr}\,\bar{f}_{\rm cr,\,0}\right)\right] = -\left[\tilde{D}_{\mu\mu}\,\bar{f}_{\rm cr,\,1} + \tilde{D}_{\mu p}\,\frac{\partial\bar{f}_{\rm cr,\,0}}{\partial p_{\rm cr}}\right]$$

$$\tilde{D}_{pp} = \chi\,\frac{p_{\rm cr}^2\,v_A^2}{v_{\rm cr}^2}\,\bar{\nu}_{\rm s}\ ,\ \tilde{D}_{p\mu} = \frac{p_{\rm cr}\,\bar{\nu}_A}{v_{\rm cr}}\,\bar{\nu}_{\rm s}\ ,\ \tilde{D}_{\mu\mu} = \bar{\nu}_{\rm s}\ ,\ \tilde{D}_{\mu p} = \chi\,\frac{p_{\rm cr}\,\bar{\nu}_A}{v_{\rm cr}}\,\bar{\nu}_{\rm s} \tag{1, 2}$$

where $\bar{f}_{\rm cr,\,n} \equiv \langle\mu^n\,f_{\rm cr}\rangle_{\mu}$ is the n'th pitch-angle moment (so e.g. $\bar{f}_{\rm cr,\,0}$ is the isotropic part of the DF, and $\bar{f}_{\rm cr,\,1} = \langle\mu\rangle\,\bar{f}_{\rm cr,\,0}$). In equations 1–2, $D_t X \equiv \partial_t X + \nabla\cdot(\mathbf{u}_{\rm gas}\,X) \equiv \rho\,{\rm d}_t(X/\rho)$ is the conservative comoving derivative (with ρ the gas mass density), $v_{\rm cr} = \beta_{\rm cr}\,c$ is the CR velocity, $p = \gamma_{\rm cr}\,\beta_{\rm cr}\,m_{\rm cr}\,c$ the CR momentum, $\hat{\mathbf{b}} \equiv \mathbf{B}/\lvert\mathbf{B}\rvert$ the unit magnetic field vector, $j_{\rm cr}$ represents injection & catastrophic losses, $\mathcal{R}_{\rm loss}$ represents continuous loss processes, v_A is the Alfvén speed, the coefficients $\tilde{D}$ are defined in terms of the scattering rate $\bar{\nu}_{\rm s} \equiv \bar{\nu}_{\rm s,+} + \bar{\nu}_{\rm s,-}$ (the scattering contributed by forward-and-backward propagating modes with respect to $\hat{\mathbf{b}}$), the signed $\bar{v}_A \equiv v_A\,(\bar{\nu}_{\rm s,+} - \bar{\nu}_{\rm s,-})/(\bar{\nu}_{\rm s,+} + \bar{\nu}_{\rm s,-})$, and the Eddington tensor $\mathbb{D}_{\rm cr} \equiv \chi\,\mathbb{I} + (1-3\,\chi)\,\hat{\mathbf{b}}\hat{\mathbf{b}}$ and scattering terms are defined in terms of $\chi \equiv (1 - \langle\mu^2\rangle)/2 = (1 - \bar{f}_{\rm cr,\,2}/\bar{f}_{\rm cr,\,0})/2$.

Integrating over an infinitesimal range in momentum for a CR group or 'packet,' this can be further transformed into the differential CR energy equation, which will be useful below:

$$D_t\,e'_{\rm cr} + \nabla\cdot\left(F'_{e,\,\rm cr}\,\hat{\mathbf{b}}\right) = \tilde{\mathcal{S}}'_{\rm sc} - \mathbb{P}'_{\rm cr}:\nabla\mathbf{u}_{\rm gas} + \mathcal{S}'_{\rm other,\,cr}$$

$$D_t\,F'_{e,\,\rm cr} + c^2\,\hat{\mathbf{b}}\cdot\left(\nabla\cdot\mathbb{P}'_{\rm cr}\right) = -\bar{\nu}_{\rm s}\left[F'_{e,\,\rm cr} - 3\,\chi\,\bar{v}_A\,\left(e'_{\rm cr} + P'_{\rm cr}\right)\right] \tag{3}$$

where $e'_{\rm cr} \equiv de_{\rm cr}/d\ln p_{\rm cr} = p_{\rm cr}^3\int d\mu\int d\phi\,E(p_{\rm cr})\,f$ is the total CR energy in a differential range of momentum $p_{\rm cr}$, $F'_{e,\,\rm cr} \equiv dF_{e,\,\rm cr}/d\ln p_{\rm cr} = p_{\rm cr}^3\int d\mu\int d\phi\,E(p_{\rm cr})\,v\,\mu\,f$ is its flux, $\mathcal{S}'_{\rm other,\,cr}$ collects any other arbitrary sources/sinks (e.g. catastrophic losses, injection at shocks, etc.), $\mathbb{P}'_{\rm cr} \equiv 3\,P'_{\rm cr}\,\mathbb{D}_{\rm cr}$ (with $P'_{\rm cr} \equiv \beta_{\rm cr}^2\,e'_{\rm cr}/3$), and scattering gives rise to the energy loss/gain term[4] $\tilde{\mathcal{S}}'_{\rm sc} \equiv -(\bar{\nu}_{\rm s}/c^2)\,\left[\bar{v}_A\,F'_{e,\,\rm cr} - 3\,\chi\,v_A^2\,\left(e'_{\rm cr} + P'_{\rm cr}\right)\right]$.

2.3 Which Alfvén speed?

In partially ionized gas, the Alfvén speed v_A is not wavelength-independent. The correct Alfvén speed in the CR dynamics equations should be Alfvén speed of gyroresonant Alfvén waves, as the original derivation of the relevant scattering terms takes $v_A =$

[3] As shown in Section 2.8 below, equation 1 reduces to the somewhat more familiar anisotropic 'streaming + diffusion' equation for $\bar{f}_{\rm cr,0}$ if one assumes $D_t\,\bar{f}_{\rm cr,1}$ is small (i.e. the flux is in 'local steady-state').

[4] As shown in Hopkins et al. (2022b) the $\tilde{D}_{\mu p}$ and $\tilde{D}_{pp}$ terms, and corresponding dependence on $\partial/\partial p_{\rm cr}$, do appear implicitly to leading-order in $\mathcal{O}(u/c)$ in equation 3 in the $\tilde{\mathcal{S}}'_{\rm sc}$ term, contributing to the 'streaming loss' and 'diffusive reacceleration' portions of $\tilde{\mathcal{S}}'_{\rm sc}$, respectively.

$v_{A,\,\rm eff} \equiv {\rm Real}(\omega_A/k_\parallel)$ with $k_\parallel = 1/\mu\, r_{g,\rm cr}$ (for gyroradius $r_{g,\rm cr}$; see e.g. Skilling 1975a). In the MHD limit (assuming e.g. non-relativistic electron gyroradii are vanishingly small), the relevant dispersion relation is:

$$\omega_A^3 + i\,\omega_A^2\,\nu_{\rm in}\,(1+\psi_{\rm ion}) - k_\parallel^2\,v_{A,\,\rm ion}^2\,\omega_A - i\,k_\parallel^2\,v_{A,\,\rm ion}^2\,\psi_{\rm ion}\,\nu_{\rm in} = 0, \quad (4)$$

where $\nu_{\rm in}$ is the ion-neutral collision frequency (this is distinct from $\nu_{\rm ni} \equiv \psi_{\rm ion}\,\nu_{\rm in}$), $v_{A,\,\rm ion} \equiv (B^2/4\pi\rho_{\rm ion})^{1/2}$, and $\psi_{\rm ion} \equiv \rho_{\rm ion}/\rho_{\rm neutral} \equiv f_{\rm ion}/(1-f_{\rm ion})$, in terms of the ion and neutral mass densities $\rho_{\rm ion}$, $\rho_{\rm neutral}$ ($\rho = \rho_{\rm ion} + \rho_{\rm neutral}$). The exact solutions to this are quite cumbersome, but the real part of interest can be well-approximated in all relevant limits by:

$$v_{A,\,\rm eff}^2 \approx v_{A,\,\rm ideal}^2\left[1+\frac{1}{\psi_{\rm ion}\,[1+\psi_{\rm ion}\,(\psi_{\rm ion}+1/4)\,\tilde{\psi}_{\rm in}^2]}\right]$$
$$\tilde{\psi}_{\rm in} \equiv \frac{\nu_{\rm in}}{k_\parallel\,v_{A,\,\rm ideal}} \approx 0.01\,(1-f_{\rm ion})\,\rho_{-24}^{3/2}\,R_{\rm GV}\,T_{1000}^{1/2}\,B_{\mu\rm G}^{-2}\,\left(k_\parallel\,r_{g,\rm cr}\right)^{-1} \quad (5)$$

where $v_{A,\,\rm ideal} \equiv (B^2/4\pi\,\rho)^{1/2}$, $f_{\rm ion}$ is the ionized fraction, $T_{1000} \equiv T/1000\,{\rm K}$, $B_{\mu\rm G} \equiv |\mathbf{B}|/\mu{\rm G}$, $\rho_{-24} \equiv \rho/10^{-24}\,{\rm g\,cm^{-3}}$, and $R_{\rm GV} \equiv R_{\rm cr}/{\rm GV}$. This essentially interpolates between the 'ideal MHD Alfvén speed' $v_{A,\,\rm ideal}$ for long-wavelength modes with frequencies ω_A much lower than the ion-neutral collision frequency $\nu_{\rm in}$, and the 'ion Alfvén speed' $v_{A,\,\rm ion}$ for short-wavelength modes with $\omega_A \gg \nu_{\rm in}$. For most ISM conditions at the (short) gyroresonant wavelengths of interest for CR dynamics, $\omega_A \gg \nu_{\rm in}$.

2.4 Scattering rates of CRs from a population of magnetic fluctuations

Everything needed to evolve the CRs in equations (1–2) is determined by the local plasma properties, except for the scattering rates $\bar{\nu}_{s,\pm}$, which crucially determine how CRs propagate. Following Zweibel (2013, 2017), quasi-linear theory gives the scattering coefficients:

$$\nu_{s,\pm}(p_{\rm cr},\,\mu) = \frac{\pi}{4}\,\Omega_{\rm cr}\,\frac{k_\parallel\,\mathcal{E}_\pm(k_\parallel)}{e_{\rm B}} \rightarrow \bar{\nu}_{s,\pm} \equiv \frac{\pi}{4}\,\hat{\nu}_s\,\Omega_{\rm cr}\,\frac{e_\pm}{e_{\rm B}} \quad (6)$$

where $e_{\rm B} \equiv |\mathbf{B}|^2/8\pi$, $k_\parallel = \Omega_{\rm cr}/(\mu\,v_{\rm cr})$ from the gyroresonant condition, and $e_\pm \equiv k_\parallel\,\mathcal{E}_\pm(\mathbf{k}\cdot\hat{\mathbf{b}} = \pm k_\parallel)$ is the energy of scattering waves at parallel wavenumber $k_\parallel$ (it is important here that we distinguish $k_\parallel \equiv |\mathbf{k}\cdot\hat{\mathbf{b}}|$ from $k = |\mathbf{k}|$). We parametrize our ignorance of the pitch-angle dependence with $\bar{\nu}_{s,\pm} \equiv (\pi/4)\,\hat{\nu}_s\,\Omega_{\rm cr}\,e_\pm/e_{\rm B}$, where $\hat{\nu}_s\,e_\pm$ reflects the angle-averaged energy of wave-packets that interact significantly with the relevant CRs (with $\hat{\nu}_s$ a dimensionless order-unity constant coming from the integration over pitch angle), traveling in either the forward (+) or backward (−) direction along $\hat{\mathbf{b}}$. As shown in Zweibel (2013, 2017) and later in Thomas & Pfrommer (2019), one can write a fluid equation for the wavepackets:

$$D_t e_\pm + \nabla\cdot\left(v_{A,\pm}\,e_\pm\,\hat{\mathbf{b}}\right) = -\tfrac{e_\pm}{2}\,\nabla\cdot\mathbf{u}_{\rm gas} + S_\pm - Q_\pm \quad (7)$$

where one can think of $e_\pm/2$ as the 'pressure' or 'PdV' term (with $\nabla\cdot\mathbf{u}_{\rm gas}$ being the change of comoving volume), we define $v_{A,\pm} = \pm v_{A,\,\rm eff}$ corresponding to the $e_\pm$ sign, and $S_\pm$ and $Q_\pm$ correspond to source and damping terms. We can write

$$S_\pm \equiv S_{{\rm sc},\pm} + S_{{\rm et},\pm} + S_{{\rm new},\pm}$$
$$Q_\pm \equiv \Gamma_\pm\,e_\pm = \left[\Gamma_{\rm in} + \Gamma_{\rm dust} + \Gamma_{\rm turb/LL} + \Gamma_{{\rm nll},\pm} + \Gamma_{{\rm new,\,damp},\pm}\right]\,e_\pm. \quad (8)$$

Here, $S_{{\rm sc},\pm}$ corresponds to energy transfer from the CRs themselves as they scatter off the waves, $S_{{\rm et},\pm}$ corresponds to 'ET' driving (defined below), and $S_{{\rm new},\pm}$ corresponds to some other, new source(s) of driving that we will consider later. Likewise, $\Gamma_\pm$ is an effective damping rate, which we will take in general to be the sum of ion-neutral ($\Gamma_{\rm in}$), dust ($\Gamma_{\rm dust}$), linear Landau or 'turbulent' ($\Gamma_{\rm turb/LL}$), non-linear Landau (NLL; $\Gamma_{{\rm nll},\pm}$), and some other arbitrary additional ($\Gamma_{{\rm new,\,damp},\pm}$) damping rates (all defined below).

From equation (3), note following Hopkins et al. (2022b) that the $\bar{\mathcal{S}}'_{\rm sc}$ term arises directly from taking the moments of the quasi-linear theory scattering rate equations for any scattering rate expressions $\nu_s(\mu,\,\mathbf{p}_{\rm cr},\,\ldots)$: it is the total energy exchange between CRs and the 'scatterers.' If we assume gyroresonance, so that CRs of a given $R_{\rm cr}$ interact only with the gyroresonant wavepacket,[5] then without making any specific assumptions about the mechanism for this exchange, energy conservation *imposes* the form of $S_{{\rm sc},\pm}$:

$$S_{{\rm sc},\pm} = \textstyle\sum_{\rm species}\,\bar{\nu}_{s,\pm}\,\frac{v_{A,\pm}}{c^2}\,\left[F'_{e,\,\rm cr} - v_{A,\pm}\,3\,\chi\,\left(e'_{\rm cr} + P'_{\rm cr}\right)\right] \quad (9)$$

Here, $\sum_{\rm species}$ represents the sum over all CR species with a given gyroresonant wavelength/rigidity $k_\parallel \sim 1/r_{g,\,\rm cr} \propto 1/R_{\rm cr}$.

2.5 Damping of parallel alfvén waves: standard mechanisms

We stress that there are many known damping processes contributing to $Q_\pm = \Gamma_\pm\,e_\pm$ for the relevant high-frequency scattering modes. Here, we briefly review a few that are commonly invoked, focusing on the terms which apply to weakly damped Alfvén modes (as compared to fast magnetosonic modes, which are vastly more strongly damped on gyroresonant scales, a case we discuss in Section 2.7 and Appendix C below).

(i) Ion-neutral damping: ion-neutral collisions generically lead to a damping rate in equation (4), which is rather complicated but for all limits where it is relevant can be accurately approximated as $Q_{{\rm in},\pm} \equiv \Gamma_{\rm in}\,e_\pm$ with $\Gamma_{\rm in} \approx (\alpha_{\rm iH} + \alpha_{\rm iHe})/2\,\rho_{\rm i} \approx 10^{-9}\,{\rm s^{-1}}\,f_{\rm neutral}\,(T/1000\,{\rm K})^{1/2}\,(\rho/10^{-24}\,{\rm g\,cm^{-3}})$.

(ii) Dust damping: from Squire et al. (2021), charged dust will have gyro motion excited by Alfvén waves on the wavelengths of interest, removing some of the scattering-wave energy (and dissipating it with dust collisions with ions + neutrals), giving a damping term: $Q_{{\rm dust},\pm} \equiv \Gamma_{\rm dust}\,e_\pm$ with $\Gamma_{\rm dust} \approx 0.02\,k\,v_{A,\,\rm eff}\,f_{\rm dg}\,(k/k_d)^{-\xi_d}$ where $f_{\rm dg} \approx 0.01\,Z/Z_\odot$ is the dust-to-gas ratio [normalized to the local ISM (LISM) value], and $k_d \sim 7.4\times10^{-11}\,{\rm cm^{-1}}\,(n/{\rm cm^{-3}})^{1/2}\,\psi_d$ with $\xi_d = 1/4$ at $k \ll k_d$ and $\xi_d = 1/2$ at $k \gg k_d$, and $\psi_d \equiv \bar{U}_0\,(\bar{\rho}_d/{\rm g\,cm^{-3}})\,(T/10^4\,{\rm K})^{\xi_{d,\,T}} \sim 1$ depending on $\bar{U}_0$ and $\bar{\rho}_d$, which parametrize the grain charge and internal grain density, $\xi_{d,\,T} = 0-1$ depending on the wavelength and grain charge regime (see equation 18 therein).

(iii) NLL damping: on gyroresonant scales, oblique magnetosonic waves are rapidly damped by resonant ion interactions, pressure anisotropy, and other effects (Lee & Völk 1973; Foote & Kulsrud 1979; Cesarsky & Kulsrud 1981; Völk & Cesarsky 1982; Squire, Quataert & Schekochihin 2016; Squire, Schekochihin & Quataert 2017). This strongly suppresses isotropic magnetosonic modes at the scales we follow, as noted above. But even for the weakly

[5] More generally as noted above, at a given momentum CRs can resonate with short-wavelength modes $k_\parallel = \Omega_{\rm cr}/(\mu\,v_{\rm cr}$ which depends on particle pitch angle, so this should be taken to be some effective pitch-angle-average over different wavenumbers. For a close-to-isotropic CR DF (as required by observations), it is straightforward (albeit tedious) to show this does not change any of our conclusions. A more detailed calculation (e.g. Kempsi et al., private communication) shows the same even for anisotropic DFs, if the dependence of scattering rate on energy predicted by SC theory $\delta_s \approx 0$ (as we show below).

damped waves of interest (e.g. parallel Alfvén modes), wave–wave interactions and field-line wandering transfer energy from the weakly damped modes to strongly damped modes, giving rise to the usual NLL damping expression $Q_{\rm nll,\pm} \equiv \Gamma_{\rm nll,\pm}\, e_\pm \equiv \Gamma^0_{\rm nll}\,(e_\pm/e_{\rm B})\, e_\pm$ with $\Gamma^0_{\rm nll} \approx (\sqrt{\pi}/8)\, c_s\, k$ (Kulsrud & Pearce 1969; Volk & McKenzie 1981).

(iv) 'Turbulent,' linear Landau, and collisionless damping: in addition to NLL damping, if there are other modes present as part of an extrinsic turbulent cascade (in addition to the parallel scattering modes themselves), these will also contribute to shearing apart or mixing waves such that power is transferred either to (a) a weakly damped but higher-k modes ('turbulent damping'; Yan & Lazarian 2002; Farmer & Goldreich 2004), or (b) strongly damped magnetosonic modes ('linear Landau damping'; Zweibel 2017). This gives $Q_{\rm turb/LL,\pm} \equiv \Gamma_{\rm turb/LL}\, e_\pm$ with $\Gamma_{\rm turb/LL}(\mathbf{k}) \sim 1/\tau_{\rm cas}(\mathbf{k})$ scaling with the cascade time-scale $\tau_{\rm cas}$ for modes of the given $\mathbf{k}$. Considering a realistically anisotropic Goldreich & Sridhar (1995)-type (GS95) cascade (which is not strongly damped on gyroresonant scales), interacting with primarily parallel modes, gives $\Gamma_{\rm turb/LL} = \Gamma_{\rm turb/LL,\,GS95} \sim [(v_{A,\,\rm ideal} + 0.4\, c_s)/\ell_A]\,(k_\parallel\, \ell_A)^{1/2}$ (where the $v_{A,\,\rm ideal}$ and c_s terms represent 'turbulent' and 'linear Landau', respectively from Farmer & Goldreich 2004; Zweibel 2017).[6]

As discussed above, it is instructive to write the damping terms in the generic form $Q_\pm \equiv \Gamma_\pm\, e_\pm$ with $\Gamma_\pm = k_\parallel^{\xi_k}\, e_\pm^{\xi_A}\, \epsilon_{\rm cr}^{\xi_{\rm cr}}\, f^\Gamma_{\rm ISM}$, where $f^\Gamma_{\rm ISM} \equiv f^\Gamma_{\rm ISM}(\rho,\ T,\ f_{\rm neutral},\ |\mathbf{B}|,\ f_{\rm dg},\ \ldots)$ is a function of 'bulk' ISM plasma properties (which do not directly depend on the CRs or $\mathbf{k}$ or $e_\pm$), and we have separated the dependence on the CR kinetic energy density $\epsilon_{\rm cr}$, scattering-wave energy $e_\pm$, and wavelength $k_\parallel$. All the damping mechanisms above give $\xi_{\rm cr} = 0$ with $0 \le \xi_A \le 1$ and $0 \le \xi_k \le 1$, as summarized for reference in Table 2.

2.6 'Self-confinement' driving

The standard 'SC' limit arises if $S_\pm \rightarrow S_{\rm sc,\pm}$ (or $S_{\rm et,\pm}$, $S_{\rm new,\pm} \rightarrow 0$), i.e. the only source term for $e_\pm$ is the CR scattering itself. This excites parallel Alfvén modes, which then compete against the different damping mechanisms in Section 2.5 to set $e_\pm$. As shown below, if the CR flux ($D_t F'_{e,\rm cr}$) and $D_t e_\pm$ equations reach local steady-state, then one of $e_\pm$ is damped to negligible values while the other (opposing the direction of the CR flux) becomes large, and the salient driving term becomes $S_{\rm sc} \rightarrow -v_{A,\,\rm eff}\, \hat{\mathbf{b}}\cdot\nabla P'_{\rm cr}$.

Akin to the damping terms, it is useful to parametrize this in terms of $S_{\rm eff} = k_\parallel^{\zeta_k}\, e_\pm^{\zeta_A}\, \epsilon_{\rm cr}^{\zeta_{\rm cr}}\, f^S_{\rm ISM}$, where $f^S_{\rm ISM}$ parametrizes the ISM structure dependence. In local flux-steady-state, the SC limit therefore gives $\zeta_k = 0$, $\zeta_A = 0$, $\zeta_{\rm cr} \approx 1$, as denoted in Table 2.[7] But this is not the only possible source of $e_\pm$ – other drivers can be included as we discuss below.

2.7 'Extrinsic turbulence' driving

In the classic ET picture, the source term for $e_\pm$ is dominated by an external turbulent cascade from much larger scales, which we will denote $S_{\rm et,\pm}$ (with symmetric $S_{\rm et,+} \approx S_{\rm et,-} = S_{\rm et}$). The traditional K41 scenario, for example, is immediately recovered if we assume an isotropic, un-damped (except for cascade transfer), inertial-range cascade, so $S_{\rm et,\pm} \rightarrow S_{\rm et,\,K41} \sim |\delta v^3|\, k \sim$ constant is just the turbulent dissipation/cascade rate, balanced by the damping (cascade transfer) term defined in Section 2.5 above $Q_\pm \rightarrow \Gamma_{\rm turb,\,K41}\, e_\pm$ with $\Gamma_{\rm turb,\,K41} \sim k\,|\delta v|$, so we obtain $e_\pm \approx e_{\rm B}\,(k\,\ell_A)^{-2/3}$, with $k_\parallel \sim k$. Similarly one could in principle imagine models that might give different isotropic power spectra such as $e_\pm \sim e_{\rm B}\,(k\,\ell_A)^{-1/2}$ (often called 'Kraichnan' or 'IK-like' in the CR literature). But as noted above (Section 2.1), and reviewed in more detail in Appendix C, these scalings cannot physically apply at gyroresonant scales for CRs with rigidities $\lesssim 0.1-1$ TV, far smaller than the Alfvén and magnetosonic dissipation scales of turbulence.

First, consider an Alfvénic (or slow-mode) cascade. It is well-established that an Alfvénic cascade cannot be isotropic on scales smaller than the Alfvén scale ℓ_A, and recall $r_{g,\rm cr} \ll \ell_A$ by a huge factor. In any cascade in which the anisotropy obeys some kind of critical balance-type condition (as seen in the solar wind, Chen 2016, and essentially all simulations of MHD turbulence cascades; see e.g. Sridhar & Goldreich 1994; Goldreich & Sridhar 1995; Boldyrev 2006; Terry 2018; Beresnyak 2019; Schekochihin 2022, and references therein), the cascade power spectrum as a function of the *parallel* wavenumber $k_\parallel$ must obey $\mathcal{E}(k_\parallel) \propto k_\parallel^{-2}$ (where critical balance gives $k_\parallel \sim |\delta v(k_\perp)/v_{A,\,\rm ideal}|\, k_\perp \ll k_\perp$, so $|k_\perp| \sim k$ and this is independent of the form of $\mathcal{E}(k_\perp)$). In other words, regardless of the dominant structure of the cascade, $e_\pm = e_{\rm B}\,\alpha_t(k_\parallel)\,(k_\parallel\,\ell_A)^{-1}$ and[8] $S_{\rm et,\pm} = \alpha_t(k_\parallel)\, e_{\rm B}\,\Gamma_{\rm turb}\,(k_\parallel\,\ell_A)^{-1}$ where more careful calculation gives the dimensionless pre-factor $\alpha_t(k_\parallel) \approx 7\,(v_{A,\,\rm eff}/v)\,\ln(k_\parallel\,\ell_A) \ll 1$ as a geometric factor that accounts for gyro-averaging over the modes with $k_\perp \gg k_\parallel$ (Chandran 2000). We further show in Appendix C that any mathematically consistent Alfvénic cascade, even one that does not follow critical balance, must obey a similar constraint on $\mathcal{E}(k_\parallel)$.

Alternatively, consider a fast-magnetosonic-mode cascade, which at least in principle could be isotropic. But recall, $r_{g,\rm cr}$ is well below the relevant dissipation/Kolmogorov scale ($k_{\rm diss} \sim 1/\ell_{\rm diss}$) of the magnetosonic cascade, so we cannot extrapolate an un-damped isotropic magnetosonic cascade from large scales. For strictly gyroresonant interactions, this generally leads to a strong suppression of the power $S_{\rm et,\pm}$ on scales $r_{g,\rm cr} \ll \ell_{\rm diss}$. Yan & Lazarian (2004) argue that if the 'resonance function' is strongly broadened by super-Alfvénic turbulence on large scales (of order the CR mean-free-path), then under the right conditions (plasma $\beta_{\rm plasma} < 1$ or $v_{A,\,\rm ideal} > c_s$, and negligible ion-neutral damping or $f_{\rm neutral} \lesssim 0.001\,(v^2_{A,\,\rm ideal}/n_1\, c_s^2)^{3/4}(c_s/10\,{\rm km\,s^{-1}})^{1/4}({\rm kpc}/R_{\rm GV}\,\ell_A)^{1/2}$), a substantial contribution to the CR scattering rate can come from the transit-time-damping terms owing to magnetosonic modes with $k \sim k_{\rm diss} \ll 1/r_{g,\rm cr}$. In that limit, the resulting scattering rates

[6] Lazarian (2016) note that on scales $1/k_\parallel$ approaching or larger than the driving and Alfvén scales this could steepen to $\Gamma_{\rm turb/LL} \propto k_\parallel^{2/3}$, but that is well outside the relevant range of scales for $\lesssim$ TeV CRs (although assuming such a scaling has no effect on our conclusions). Likewise, the Farmer & Goldreich (2004) argument that field-line fluctuations set a minimum $|k_\perp/k_\parallel| \sim |\delta B_{\rm ext}(k_\perp)|/|B_0|$ can, in principle, be generalized for any external critically balanced shearing cascade with different intermittency effects modifying the perpendicular spectrum (Schekochihin 2022), but these generally lead to only minor modifications of $\Gamma_{\rm turb/LL} \sim [(v_{A,\,\rm ideal} + 0.4\, c_s)/\ell_A]\,(k_\parallel\,\ell_A)^{\xi_k}$ with $0.4 \lesssim \xi_k \lesssim 0.5$.

[7] When the CR flux and $e_\pm$ equations are far from local quasi-equilibrium, then from equation (9) the more general form of $S_{\rm cr}$ ($\propto \bar{v}_{\rm s}\,(v_{A,\pm}/c)\,[F'_{e,\rm cr} - v_{A,\pm}{}^3\chi\,(e'_{\rm cr} + P'_{\rm cr}) \propto \Omega_{\rm cr}\,(e_A/e_{\rm B})\,\epsilon_{\rm cr} \propto k\, e_A\, \epsilon_{\rm cr}$) would have coefficients closer to $\zeta_k \sim 1$, $\zeta_A \sim 1$, $\zeta_{\rm cr} \sim 1$. But because these reach equilibrium on short time-scales, and both expressions have $\zeta_{\rm cr} \sim 1$, which is the important feature that drives the qualitative behaviour discussed below, in our analytic models, we will typically work with the local steady-state expressions.

[8] The statement $S_{\rm et} \propto \Gamma_{\rm turb}$ here is just a rephrasing of the usual relation between the turbulent power spectrum and the cascade rate $\tau_{\rm cas} \sim 1/\Gamma_{\rm turb}$, i.e. $e_A \sim k_\parallel\,\mathcal{E}(k_\parallel) \propto S(k_\parallel)/\,\Gamma_{\rm turb}(k_\parallel) \sim S(k_\parallel)\,\tau_{\rm cas}(k_\parallel)$.

from the Yan & Lazarian (2004) model can be written as $\bar{\nu}_s \sim \Omega_{cr}\, e_B^{-1}\, (\delta B^2[k_{diss}])\, (k_{diss}/k_\parallel)$ (Hopkins et al. 2021d). Even though by definition in this scenario some of the scattering modes come from larger scales, in our mathematical formalism, this is identical to assuming an equivalent gyroresonant mode or cascade power with $e_\pm \sim (\delta B^2[k_{diss}])\, (k_{diss}/k_\parallel)$ or (since $\delta B[k_{diss}]$ comes, by assumption in this model, from some isotropic cascade on larger scales with $|\delta B^2[k_{diss}]| \sim B^2\, (k_{diss}\, \ell_A)^{-\psi_{turb}}$) therefore $e_\pm \sim e_B\, (k_{diss}\, \ell_A)^{-(\psi_{turb}+1)}\, (k_\parallel\, \ell_A)^{-1}$. Again, we show in Appendix C that a similar constraint is generic to any magnetosonic cascade with $\ell_{diss} \gtrsim r_{g,cr}$ (independent of its detailed form or which terms dominate the CR scattering).

Parametrizing again as $S_{eff} = k_\parallel^{\zeta_k}\, e_\pm^{\zeta_A}\, \epsilon_{cr}^{\zeta_{cr}}\, f_{ISM}^{S}$, we see that in the ET limit generically $\zeta_{cr} = \zeta_A = 0$ and accounting for anisotropy and/or damping, we must have $\zeta_k \lesssim -1 + \xi_k^{turb}$ (where ξ_k^{turb} is the dependence of $\Gamma_{turb} \propto k_\parallel^{\xi_k^{turb}}$), which is equivalent to $\delta_s \lesssim 0$. For a magnetosonic ET cascade, if dissipation is non-negligible outside the special limits above (e.g. in gas with plasma $\beta_{plasma} > 1$, or partially neutral gas), this gives a super-exponential cutoff to the ET power spectrum on small scales, which is equivalent to $\zeta_k \ll -1 + \xi_k^{turb}$.

In Appendix C, we present a much more detailed review and discussion of the anisotropy and damping constraints above. There we show that the key conclusion that any cascade model must predict $\zeta_k < -1 + \xi_k^{turb}$, i.e. $e_\pm \propto k^{-1}$ or steeper if $r_{g,cr}$ is smaller than the dissipation and Alfvén scale, is robust to any specific assumptions about the turbulent cascade, dissipation mechanism, or mode structure.

2.8 Steady-state solutions

Equations (7) and (2) converge to their 'local steady state' or quasi-equilibrium values with $D_t e_\pm \rightarrow 0$ and $D_t \bar{f}_{cr,1} \rightarrow 0$ (or more formally, $|D_t e_\pm| \ll |\bar{\nu}_s\, e_\pm|$, $|D_t \bar{f}_{cr,1}| \ll |\bar{\nu}_s \bar{f}_{cr,1}|$), in approximately the scattering time $\bar{\nu}_s^{-1}$. This is $\bar{\nu}_s^{-1} \sim 30\,{\rm yr}\, R_{GV}^{0.5}$ from empirically fitted models, much faster than other time-scales over which e.g. bulk ISM properties evolve (see Appendix B). In this limit, assuming strong scattering, the DF becomes nearly isotropic ($\chi \approx 1/3$) and equations 1–2 can be combined into a single anisotropic diffusion equation:

$$D_t \bar{f}_{cr,0} \approx \nabla\cdot\left(\kappa_\parallel\, \hat{\mathbf{b}}\hat{\mathbf{b}}\cdot\nabla \bar{f}_{cr,0} + \frac{\bar{v}_A}{3}\,\hat{\mathbf{b}}\, p_{cr}\,\frac{\partial \bar{f}_{cr,0}}{\partial p_{cr}}\right) + j_{cr,0} + \frac{1}{p_{cr}^2}\frac{\partial}{\partial p_{cr}}\left[p_{cr}^3\left\{\frac{\mathcal{R}_{loss}}{p}\,\bar{f}_{cr,0} + \frac{\nabla\cdot\mathbf{u}_{gas}}{3}\,\bar{f}_{cr,0} - \frac{\bar{v}_A}{3}\,\hat{\mathbf{b}}\cdot\nabla\bar{f}_{cr,0} + \frac{\left(v_{A,\,eff}^2 - \bar{v}_A^2\right)}{9\,\kappa_\parallel}\,p\,\frac{\partial \bar{f}_{cr,0}}{\partial p}\right\}\right] \tag{10}$$

with $\kappa_\parallel \equiv v_{cr}^2/3\,\bar{\nu}_s$. The relevant behaviours here are more obvious if we again take the CR energy equation, equation (3):

$$D_t e'_{cr} \approx \nabla\cdot\left[\kappa_\parallel\,\hat{\mathbf{b}}\hat{\mathbf{b}}\cdot\nabla e'_{cr} - \bar{v}_A\,\hat{\mathbf{b}}\,e'_{cr}\right] - P'_{cr}\,\nabla\cdot\left(\mathbf{u}_{gas} + \bar{v}_A\,\hat{\mathbf{b}}\right) + \frac{v_{A,\,eff}^2 - \bar{v}_A^2}{c^2}\,\bar{\nu}_s\,\left(e'_{cr} + P'_{cr}\right) + \mathcal{S}'_{other,\,cr}, \tag{11}$$

illustrating that we have anisotropic 'diffusion' with $\kappa_\parallel \equiv v_{cr}^2/3\,\bar{\nu}_s = v_{cr}^2/[3\,(\bar{\nu}_{s,+} + \bar{\nu}_{s,-})]$, and 'streaming' along $\hat{\mathbf{b}}$ with speed $\bar{v}_A \equiv v_{A,\,eff}\,(\bar{\nu}_{s,+} - \bar{\nu}_{s,-})/(\bar{\nu}_{s,+} + \bar{\nu}_{s,-})$.

Meanwhile, equation (7) becomes $\nabla\cdot(v_{A,\pm}\,e_\pm\,\hat{\mathbf{b}}) + e_\pm\,(\nabla\cdot\mathbf{u}_{gas})/2 + Q_\pm - S_\pm \approx 0$, with $S_{sc\pm}$ as a function of F'_{cr} given by solving the steady-state flux equation $F'_{e,cr} \approx \bar{v}_A\,(e'_{cr} + P'_{cr}) - (c^2/\bar{\nu}_s)\,\hat{\mathbf{b}}\cdot\nabla P'_{cr}$. This gives $S_{sc,\pm} \approx -(\nu_{s,\pm}/\bar{\nu}_s)\,v_{A,\pm}\,\hat{\mathbf{b}}\cdot\nabla P'_{cr} - 2\,(v_A/c)^2\,(\nu_{s,\mp}/\bar{\nu}_s)\,(e'_{cr} + P'_{cr})$. Assuming we know the form of $S_{et,\pm}$, $\Gamma_\pm$, etc., the pair of steady-state equations for $e_\pm$, can then be numerically solved exactly, given a locally fixed background (**B**, etc.; see Appendix B). But the exact solution is given by a fifth-order polynomial in $e_+ + e_-$, without closed-form solutions, which is not particularly useful or instructive (defeating the purpose of our steady-state assumption here).

It is much more useful to consider two limiting cases (justified in Appendix B). First, if the extrinsic term dominates the 'source' with $|S_{et,\pm}| \gg |S_{sc,\pm}|$, then assuming the extrinsic driver is symmetric with respect to waves in the $\pm\hat{\mathbf{b}}$ direction, $S_{et,+} \approx S_{et,-}$, we have $\bar{\nu}_{s,+} \approx \bar{\nu}_{s,-}$ and $|\bar{v}_A| \ll v_A$. Alternatively, if the SC term dominates, then either $\bar{\nu}_{s,+} \gg \bar{\nu}_{s,-}$ or $\bar{\nu}_{s,+} \ll \bar{\nu}_{s,-}$ (the larger corresponding to the opposite direction of $\hat{\mathbf{b}}\cdot\nabla P'_{cr}$), so $|\bar{v}_A| \approx v_{A,\,eff}$. In either case, the 'streaming' speed $\bar{v}_A$ is not especially important for CR transport, because it is subdominant to the 'diffusive' term at most energies of interest for empirically allowed models (shown below). This can be seen both by considering its normalization, which is far smaller than allowed by empirical considerations (with a halo size ~ 10 kpc, the CR 'escape time' due to this term would be ~ 10 Gyr, compared to the observed $\sim 5\,R_{GV}^{-0.5}$ Myr), or by noting that this would produce a residence/escape time that is completely independent of rigidity, again in contradiction to CR observations. So, either in the SC or ET limit, we can reasonably reduce the key transport physics to understanding $\kappa_\parallel \propto 1/\bar{\nu}_s$, and can approximate equation (7) with a single equation for $\bar{\nu}_s$ to leading order. Moreover, for the types of models discussed above, the second and third terms in equation (7; $\sim\nabla(v_{A,\,eff}\,e_\pm)$ and $\sim e_\pm\,\nabla\mathbf{u}_{gas}$) are relatively small ($\ll S \sim Q$), so if we write $Q_\pm \equiv \Gamma_\pm\,e_\pm$, then we can very generically approximate $S \approx \Gamma_\pm\,e_\pm$ in steady-state, giving

$$\kappa_\parallel \equiv \frac{v_{cr}^2}{3\,\bar{\nu}_s} = \frac{4\,\beta_{cr}}{3\pi\,\hat{\nu}_s}\,c\,r_{g,cr}\,\left(\frac{e_B}{e_+ + e_-}\right) \sim v_{cr}\,r_{g,cr}\,\left(\frac{\Gamma_\pm\,e_B}{S_\pm}\right). \tag{12}$$

We will return to this approximate scaling below.

3 PROBLEMS OF BOTH SELF-CONFINEMENT & EXTRINSIC TURBULENCE MODELS

3.1 Empirical models

It is well-known that one can reproduce almost all of the observed local Solar neighborhood CR data and Galactic γ-ray constraints by assuming an empirically *parametrized* $\kappa_\parallel \sim \beta_{cr}\,c\,r_0(R_{cr})$..[9] Most modern studies favour a scaling close to $r_0 \sim 10^{19}\,{\rm cm}\,R_{GV}^{1/2}$, or equivalently $e_A/e_B = (e_+ + e_-)/e_B \sim 3\times10^{-7}\,B_{\mu G}^{-1}\,R_{GV}^{1/2}$, or $e_A \sim (k_\parallel\,\ell_0)^{-1/2}\,e_B$ with $\ell_0 \sim 3\times10^{25}\,B_{\mu G}^2\,{\rm cm}$ (Blasi & Amato 2012; Vladimirov et al. 2012; Gaggero et al. 2015; Cummings et al. 2016; Guo et al. 2016; Jóhannesson et al. 2016; Korsmeier & Cuoco 2016; Evoli et al. 2017; Amato & Blasi 2018). For example, De La Torre Luque et al. 2021 show that, even allowing for a wide range of model variations with different systematic uncertainties and assumptions, the largest plausible deviations in the empirical models lie within the range $r_0 \sim 10^{18.5-19.5}\,R_{GV}^{0.4-0.7}\,{\rm cm}$, i.e. $e_A \sim (10^{-7} - 10^{-6})\,\langle B_{\mu G}^{-1}\rangle\,R_{GV}^{0.3-0.6}\,e_B$, or equivalently $|\delta B(k_\parallel \sim 1/r_{g,cr})|/|\mathbf{B}| \sim (0.3-1)\,0.001\,R_{GV}^{(0.15-0.3)}$. This required scaling seems, at first, remarkably simple and plausible. Yet, in practice, it proves remarkably

[9] In models that assume an isotropic diffusion Fokker–Planck equation for CRs, it is common to quote D_{xx}. For isotropically tangled magnetic fields, $D_{xx} \approx \kappa_\parallel/3$.

difficult to actually produce even *qualitatively* similar scaling from either ET or SC models at energies $\lesssim$ TeV.

For reference below, in the CR propagation literature the slope of the dependence of diffusivity on rigidity is usually parametrized as $\kappa_\parallel \sim (\beta_{\rm cr}\, c^2/\bar{\nu}_{\rm s,\,0})\, R_{\rm GV}^{\delta_{\rm s}}$, so $\bar{\nu}_{\rm s} = \bar{\nu}_{\rm s,\,0}\, \beta_{\rm cr}\, R_{\rm GV}^{-\delta_{\rm s}}$, with $\bar{\nu}_{\rm s,\,0}$ a constant and $0.4 \lesssim \delta_{\rm s} \lesssim 0.7$ allowed by observations (above). To reproduce this, we require $e_A \propto R_{\rm cr}^{1-\delta_{\rm s}} \propto r_{g,{\rm cr}}^{1-\delta_{\rm s}} \propto k_\parallel^{\delta_{\rm s}-1}$.

3.2 The problems with extrinsic turbulence

First, let us consider the standard ET models introduced in Section 2.7. Naively, the empirically inferred slope $\delta_{\rm s}$ appears quite consistent with the expectation from an isotropic, undamped, inertial-range cascade with $\mathcal{E}(k) \propto k^{-3/2}$, which gives $e_A \sim k_\parallel\, \mathcal{E}_{\rm turb}(k_\parallel) \sim k_\parallel^{-1/2}$ and thus $\delta_{\rm s} = 1/2$). It is also marginally consistent with an isotropic undamped inertial range K41 cascade ($\delta_{\rm s} = 1/3$). But there are three major problems: (1) anisotropy, (2) damping, and (3) normalization. Once again, recall that for all CR energies of interest $r_{g,{\rm cr}} \sim 10^{12}\,{\rm cm}\, R_{\rm GV}/B_{\mu\rm G}$ is much smaller than both of the Alfvén scale – below which the turbulence is sub-Alfvénic ($\ell_A \gtrsim$ pc) – and the magnetosonic dissipation/Kolmogorov scales ($\ell_{\rm diss} \gtrsim 10^{15}$ cm).

(i) **Anisotropy:** First, (1) below the Alfvén scale, theory and simulations robustly predict something akin to critical balance must apply to the Alfvénic cascade. But as noted in Section 2.7 (see Schekochihin 2022), *any* energy-conserving cascade that obeys critical balance automatically predicts $\mathcal{E}(k) \propto k_\parallel^{-2}$ ($e_A \propto k_\parallel^{-1}$). This implies $\delta_{\rm s} = 0$, regardless of how the cascade scales with the perpendicular components of $\mathbf{k}$, because it is $k_\parallel$ that plays the key role for CR scattering.[10] But this further implies scattering rates are *independent* of CR energy, which is *strongly* ruled-out.

(ii) **Damping/Dissipation:** Second, (2) strong dissipation, particularly of magnetosonic modes, causes two problems. First, it damps fast magnetosonic fluctuations, which makes the normalization of $\bar{\nu}_{\rm s}$ much too low for scattering from a magnetosonic cascade (generally by ∼3–6 orders of magnitude for realistic damping rates, as shown in Hopkins et al. 2021d). But, equally important, dissipation can only make the spectrum steeper ($\mathcal{E}(k)$ decreases more-rapidly at high k). This in turn implies that $\delta_{\rm s}$ generically becomes *negative*. One might argue that at some point, one should ignore the strongly damped smaller-scale gyroresonant magnetosonic modes and only integrate the contribution from larger magnetosonic modes above the dissipation/Kolmogorov scale – this is the argument in Yan & Lazarian 2004 (YL04). But, as noted in Section 2.7, this *also* gives $\delta_{\rm s} \leq 0$, always. In fact $\delta_{\rm s} = 0$, i.e. $e_A \propto k_\parallel^{-1}$, corresponds to the 'most efficient' possible case of the YL04 model so long as $\ell_{\rm diss} \gg r_{g,\,{\rm cr}}$, and in many other limits $\delta_{\rm s} \ll -1$, e.g. when there is an exponential-like cutoff due to non-zero ion-neutral damping, or plasma $\beta_{\rm plasma} \equiv c_s^2/v_{A,\,\rm ideal}^2 > 1$. Again, as shown rigorously in Appendix C, this applies to *any* magnetosonic cascade regardless of details, if the gyroscale is smaller than the (fast-mode) dissipation scale. So the only way to 'salvage' even the qualitative scaling of $\delta_{\rm s}$ in ET models, with turbulence that is either Alfvénic or magnetosonic in character, is to ignore *both* anisotropy and damping/dissipation effects. This would require discarding almost everything that is known about the structure of MHD turbulence.

[10] In Appendix C, we show that $\delta_{\rm s} \leq 0$ is generic to any Alfvénic cascade, independent of its form, and mathematically allowed violations of critical balance or other conditions within a cascade generically lead to $\delta_{\rm s} < 0$, making the problem worse.

(iii) **Normalization:** Third, (3) even if we did invoke an isotropic, undamped $\mathcal{E}(k) \propto k^{-3/2}$ cascade, which leads to the observationally favoured slope, then we would obtain $e_A \sim e_{\rm B}\, (k\, \ell_A)^{-1/2}$. This is systematically larger than the empirically required value of e_A by a factor $\sim 1000\, B_{\mu\rm G}\, (\ell_A/10\,{\rm pc})^{-1/2}$ – i.e. this would typically over-predict observed CR scattering rates by factors of several thousand. Thus, some anisotropy and/or damping *must* be present to prevent ET from over-confining CRs, but as soon as those are invoked, the predicted shape ($\delta_{\rm s}$) is incorrect.

As reviewed in Hopkins et al. (2021d) and Appendix C, almost all proposed more-detailed corrections and modifications to traditional ET models in the literature make the problems above worse, not better (i.e. they make the scattering rates even more different from those observationally required).

3.3 Generic alternatives to extrinsic turbulence

A generic alternative to ET is to have scattering modes that are directly excited 'at each scale' by some process, rather than arising through a cascade from large scales. In this scenario, the driver must operate over a wide range of scales – scattering CRs in the range ∼MeV-TeV implies a factor ∼10^6 in k – and excite the relevant $k_\parallel$ modes.

Of course SC, where the excitation comes from the CRs themselves, does exactly this, and is very natural – indeed, it should occur to some extent. And a simple order-of-magnitude calculation shows that the SC source term $S_{\rm sc}$ should almost always be dominant over the 'standard' ET source terms for CR energies $\lesssim$ TeV, if we account for either anisotropy or damping of ET (let alone both). For this reason, SC has been the most popular model to explain the scaling of e_A (hence CR scattering rates) at these energies.

3.4 The problems with self-confinement

However, there are also three major qualitative problems with SC models: (1) normalization, (2) spectral shape/scaling of scattering rates, and (3) instability or 'solution collapse.'

(i) **Normalization:** The normalization problem (1) is discussed in detail in both analytic models and full dynamical simulations in Hopkins et al. (2021d). In brief, for $\sim 1-10$ GeV CRs, which contain most of the energy, while damping is large in neutral gas, the CR energy density in a multiphase ISM is determined by the volume-filling phases with the *lowest* diffusivity (the 'boundary condition') i.e. the WIM and inner CGM. In these regions, NLL and ion-neutral damping are both inefficient, so standard models suggest that turbulent or linear-Landau damping dominates (see Hopkins et al. 2021d, c; Buck et al. 2020). Considering CRs near the peak of the CR energy spectrum, and assuming local steady state and dominant turbulent damping, the effective diffusivity from equation (12) would then be (see Hopkins et al. 2021d) $\kappa_\parallel \sim v_{\rm cr}\, r_{g,\,{\rm cr}}\, (\Gamma_\pm\, e_{\rm B}/S_{{\rm sc},\pm}) \approx 10^{28}\, \delta v_{10}^{3/2}\, \ell_{\nabla,{\rm cr,kpc}}\, \ell_{A,\,10}^{-1/2}\, e_{\rm cr,\,eV}^{-1}\, n_1^{3/4}$ (with $\delta v_{10} \equiv \delta v_{\rm turb}/10\,{\rm km\,s^{-1}}$, $\ell_{\nabla,\,{\rm cr,\,kpc}} \equiv \ell_{\nabla,\,{\rm cr}}/{\rm kpc}$ with $\ell_{\nabla,{\rm cr}} \equiv P_{\rm cr}'/|\nabla P_{\rm cr}'|$, $\ell_{A,\,10} \equiv \ell_A/10\,{\rm pc}$, $e_{\rm cr,\,eV} \equiv e_{\rm cr}'/{\rm eV\,cm^{-3}}$, $n_1 \equiv n/{\rm cm^{-3}}$), or $e_A/e_{\rm B} \sim 10^{-5}\, e_{\rm cr,\,eV}\, \ell_{A,\,10}^{1/2}\, B_{\mu\rm G}^{-1}\, \delta v_{10}^{-3/2}\, \ell_{\nabla,{\rm cr,kpc}}^{-1}\, n_1^{-3/4}$. This is a factor ∼30–100 smaller than the empirically favoured value of $\kappa_\parallel$ for $\sim 1-10$ GeV CRs.

This issue is potentially important, but perhaps the least-serious of the three problems: it can be ameliorated by (a) including additional damping mechanisms such as the recently proposed dust damping, which can be larger than turbulent damping invoked above (raising $\Gamma_\pm$) by factors ∼10–100 at ~ 1 GV under MW-like conditions

(Squire et al. 2021); (b) invoking somewhat slower growth rates S, for the relevant modes as excited by CRs, which can occur accounting more accurately for e.g. the full spectrum of modes which contribute to scattering (instead of assuming strict gyroresonance plus the grey approximation), as argued in Bai et al. (2019); (c) accounting more accurately for geometric and other non-grey effects, which can lower the effective scattering rate for gyroresonant modes for a given e_A (i.e. our '$\hat{\nu}_s$' parameter). These corrections arise from accounting more accurately for the full shape of the CR spectrum (Kempski et al., in prep), accounting for the $\mu = 0$ pitch-angle scattering barrier, which has been shown to be significant in some PIC simulations (Bai et al. 2019), and accounting for anisotropy that arises from the fact that the CRs only source one helicity of modes in some regimes (Holcomb & Spitkovsky 2019). It is noteworthy that all of these corrections go in the 'favoured' direction.

(ii) **Spectral Shapes:** However problem (2), regarding the spectral-shape-dependence of scattering rates, is not resolved by the possible solutions above. The issue is that the growth term $S_{\rm sc,\pm}$ scales with the CR pressure or energy itself, as $S_{\rm sc,\pm} \rightarrow \pm v_{A,\,\rm eff}\,\hat{\mathbf{b}}\cdot\nabla P'_{\rm cr} \sim (v_{A,\,\rm eff}/\ell_{\nabla,\,\rm cr})\,\epsilon_{\rm cr}$, where $\ell_{\nabla,\,\rm cr}$ is the CR gradient scale length above and $\epsilon_{\rm cr}$ is the CR kinetic energy density in a logarithmic interval in rigidity $R_{\rm cr}$ (or equivalently, gyroresonant $k_\parallel$). But, for realistic CR spectra, $\epsilon_{\rm cr} \propto R_{\rm cr}^{-\alpha_{\rm cr}} \propto k_\parallel^{\alpha_{\rm cr}}$, with $-1.4 \lesssim \alpha_{\rm cr} \lesssim -0.7$ at energies $\ll 1\,$GV, and $\alpha_{\rm cr} \approx 0.7$ at energies from $\sim 1-300\,$GV (e.g. Cummings et al. 2016; Bisschoff et al. 2019, and references therein). So, in our parametrization above, we would have $S_{\rm sc,\pm} \propto k_\parallel^{\alpha_{\rm cr}}$, with $\Gamma_\pm \propto k_\parallel^{\xi_k}\,e_A^{\xi_A}$ (for all the known damping mechanisms discussed in Section 2.5). In steady-state, i.e. solving $S_\pm \sim \Gamma_\pm\,e_A$, this leads to $\delta_{\rm s} \approx 1-(\xi_k-\alpha_{\rm cr})/(1+\xi_A)$. But for $\alpha_{\rm cr} \sim 0.7$ (i.e. $R_{\rm cr} \gtrsim$ GV), this gives $0.8 \lesssim \delta_{\rm s} \lesssim 1.7$ for all known damping mechanisms ($0 \le \xi_k,\ \xi_A \le 1$), while for $\alpha_{\rm cr} \sim -1$ (i.e. $R_{\rm cr} \lesssim$ GV), this gives $-1 \lesssim \delta_{\rm s} \lesssim 0.5$. In other words, $\delta_{\rm s}$ is generically much too low at $R_{\rm cr} \ll 1\,$GV, owing to the fact that the spectrum is rising, and much too high at $R_{\rm cr} \gg 1\,$GV, owing to the fact that the spectrum is falling. This produces a strong minimum in $\kappa_\parallel$, a much-too-sharply peaked CR spectrum, and the incorrect dependence in both limits of e.g. secondary-to-primary ratios on $R_{\rm cr}$. In principle, this could be solved by imposing a new dominant damping mechanisms with a dependence on k or e_A that is different to the mechanisms discussed in Section 2.5. But, this would require rather unusual values of ξ_k, in particular, $\xi_k \approx \alpha_{\rm cr} + (1-\delta_{\rm s})\,(1+\xi_A)$ implies $1.2 \lesssim \xi_k \lesssim 1.7$ for $R_{\rm cr} \gtrsim 1\,$GV, and $-0.5 \lesssim \xi_k \lesssim 0$ for $R_{\rm cr} \lesssim 1\,$GV. This in turn requires a dramatic difference in the dominant damping mechanism between waves that resonate with low and high CR energies – longer wavelength waves require larger ξ_k, while shorter wavelength waves require much smaller (negative) ξ_k – compared to the known mechanisms described in Section 2.5. In other words, it is not possible to resolve this problem by simply 'tweaking' coefficients or power-law scalings of standard damping or growth-rate terms.

(iii) **Instability:** Even if (1) and (2) were solved with something like the possibilities mentioned above, a potentially more fundamental problem with self-confinement is that the commonly adopted 'local steady-state' SC solutions for the CR scattering rates and fluxes (given some CR energy density at a given CR momentum/rigidity) are not stable equilibria of the CR energy equation. This is derived in detail in Appendix A. Briefly, the problem arises because the SC driving/growth rate $S_{\rm sc,\pm}$ is proportional to the CR flux $F'_{\rm e,\,cr}$ or (in local steady-state) to $\propto \nabla P'_{\rm cr}$ (i.e. $\zeta_{\rm cr} \sim 1$), while all standard damping processes are independent of $e'_{\rm cr}$ (i.e. $\xi_{\rm cr} = 0$). This implies $e_A \sim S_{\rm sc,\pm}/\,\Gamma_\pm \propto \nabla e'_{\rm cr}$, where we assumed linear damping dominates for simplicity (this assumption is relaxed in Appendix A and our simulation tests). But this in turn means that the steady-state CR flux $F'_{e\,\rm cr} \propto \kappa_\parallel\,\nabla e'_{\rm cr} \propto e_A^{-1}\,\nabla e'_{\rm cr} \propto$ constant is *independent* of $e'_{\rm cr}$; or equivalently that the CR 'escape time' (the streaming/diffusion time to some distance ℓ) is $t_{\rm esc} \sim \ell^2/\kappa_\parallel \sim \ell/v_{\rm stream,\,eff} \propto \ell^2\,e_A/(v_{\rm cr}\,r_{g,\rm cr}\,e_{\rm B}) \propto e'_{\rm cr}$. This is a fundamental feature of SC models, as it is a re-statement of the fact that the confining wave energy ultimately comes from the CRs themselves. Indeed, the effect clearly manifests in the early kinetic quasi-linear theory and solutions of Skilling (1971), as the absence of the CR distribution in the diffusion term (the final term in their eq. (9)). The consequence of this feature of the SC model is that if the CR flux is lower than the exact value needed to maintain global steady-state in the CR energy density equation (or even if there is a perturbation about the exact value), the local $e'_{\rm cr}$ will increase due to the CR overconfinement, then causing e_A and the CR escape time to also increase, which in turn further bottlenecks[11] $e'_{\rm cr}$. Since the steady-state CR energy density in some region $e'_{\rm cr} \sim j_{\rm inj}\,t_{\rm esc}$, and both $j_{\rm inj}$ and $t_{\rm esc}$ have just increased, $e'_{\rm cr}$ increases as well. This will grow on a time-scale of the injection time ($e'_{\rm cr}/j_{\rm inj}$), until $t_{\rm esc} \propto e'_{\rm cr}$ becomes so large that the CRs can only move at the local Alfvén speed (effectively $\bar{\nu}_{\rm s} \rightarrow \infty$), which is much slower than relevant loss times. They thus lose all their energy to catastrophic or radiative losses before they can escape. As a result, in steady-state, the energy density is no longer set by the escape time but by the calorimetric loss time-scale $e'_{\rm cr} \sim j_{\rm inj}\,t_{\rm loss}$. The outcome is that CRs at all $R_{\rm cr}$ lose all their energy near their injection sites, in gross contradiction to observations. Even if we neglect losses, the CRs would all stream at the same speed v_A, implying $\delta_{\rm s} = 0$, which is also in direct contradiction to observations.

This is the 'SC runaway' problem described and seen in the simulations of Hopkins et al. (2021d), who noted it for just $\sim 1\,$GV CRs. Here, we note that it applies at all energy scales, becoming more severe due to the additional constraints on $\delta_{\rm s}$. Conversely, if one deviates from a balanced initial condition (ICs) by lowering $e'_{\rm cr}$, then e_A drops, the escape time decreases, so $e'_{\rm cr}$ drops more rapidly, etc. until CRs essentially 'free stream' and produce negligible secondaries.

We will show that, in simulations, this instability or 'solution collapse' problem is quite severe. In controlled restarts from otherwise identical ICs, if we start from a 'high' CR density $e'_{\rm cr}$ (with only SC-motivated S), then $e'_{\rm cr}$ and scattering rates grow until they reach calorimetric losses across a wide range of energies; conversely, if we start from a 'low' $e'_{\rm cr}$, then $e'_{\rm cr}$ and scattering rates decrease to an extremely low value, giving negligible secondary production (much too-low B/C, $\bar{p}/p$, e^+/e^-) at all energies. It is inevitable that real galaxies will almost constantly undergo events that push them towards one or the other limit.

3.5 How to rescue things?

We now consider what would be qualitatively needed in either the damping ($\Gamma_\pm$) or driving ($S_\pm$) terms for e_A, in order to obtain some plausible consistency between observations and either SC or ET CR transport models.

[11] In the literature, CR 'bottlenecks' can occur under a variety of different conditions and the term is often used to refer to distinct physical processes ((see e.g. Zweibel 2017; Bustard & Zweibel 2021; Huang & Davis 2022; Quataert, Thompson & Jiang 2022a). Unless otherwise specified, in this paper we will generally use the term in connection to this SC 'solution collapse' problem, specifically to the over-confined solution branch.

3.5.1 Alternative damping

First, let us consider the consequences of modifying or adding only damping terms, while retaining standard SC & ET driving.

In the ET-dominated limit – or any model where the dominant contribution to the driving term $S_{\pm}$ derives from an MHD cascade – no addition or modification of the form of the damping terms can 'rescue' the models. This is because, as shown above, the cascade assumption itself implies $\delta_s < 0$ (i.e. the slope of the 'cascade' is cut off to some non-zero degree) for *any* finite damping term if the damping is significant. Even for negligible damping in ET, anisotropy requires $\delta_s \leq 0$, independent of the form of the damping terms. However, for standard scattering expressions motivated by physical turbulence models (i.e. including anisotropy and damping), our naive expectation is that the driving term $S_{\pm}$ is dominated by SC at all energies of interest (i.e. $S_{et,\pm} \ll S_{sc,\pm}$). Thus, it is most relevant to consider what (if any) damping terms could resolve the problems above for self-confined CRs.

In the SC-dominated limit, one could in principle solve both the 'normalization' (1) and 'spectral shape' (2) problems by invoking some arbitrary damping rate $\Gamma_{new,\,damp,\pm}$, which has the appropriate normalization and desired scaling as a function of k and e_A (i.e. varying ξ_k as a function of R_{cr}, to new values outside the range of known damping mechanisms, as explained in Section 3.4). This in itself would not solve the 'instability' problem (3). However, if the dominant damping mechanism *also* scaled with e'_{cr}, then in equilibrium ($S_{\pm} \sim Q_{\pm}$) this would cancel the e'_{cr} dependence of $S_{sc,\pm}$, resolving issue (3) as well. Specifically, following the parametrization of Section 2.5, if we had damping with $\xi_k \approx 1/2$, $\xi_A \approx 0$, $\xi_{cr} \approx 1$, i.e. $\Gamma_{new,\,damp,\pm} \propto k_{\parallel}^{1/2}\, e'_{cr}$, then since the SC driving term $S_{sc,\pm} \propto v_{A,\,eff}\, \hat{\mathbf{b}} \cdot \nabla P'_{cr}$, this would give approximately the desired $e_A \propto k_{\parallel}^{-1/2}$ independent of e'_{cr}, thus curing the 'spectral shape' and 'instability' problems. An example that gives roughly the correct normalization as well would be something like $\Gamma_{new,\,damp,\pm} \sim v_{eff}\,(k_{\parallel}\,\ell_A)^{1/2}\,(e'_{cr}/e_B)$, with $v_{eff} \sim c_s \sim v_{A,\,ideal}$ – i.e. something akin to standard turbulent or linear Landau damping rates, but multiplied by e'_{cr}/e_B.

This is perhaps not wildly implausible, if it arose from some non-linear process involving CR back-reaction on the gas (perhaps, for example, shocks induced on small-scales by CR-gas coupling as in Huang & Davis 2022; Hin Navin Tsung, Oh & Jiang 2022; Quataert, Jiang & Thompson 2022b). Our simulations below will test both 'microscopic' (i.e. subgrid, added manually) and 'macroscopic' (spatially resolved, on $\gtrsim$pc scales, which should emerge self-consistently from the simulation physics) versions of this scenario. However, it is not a perfect solution. Not only would one need to think of an effective damping mechanism that produced the desired scaling above (for which there is no obvious candidate, and our simulations do not appear to produce this 'macroscopically'), but one would also need to ensure that this is the *dominant* damping process compared to the other known damping terms in Section 2.5, which do not simply disappear. This would need to be true, at least on average in the ISM, for $0.01\,\mathrm{GV} \lesssim R_{cr} \lesssim 1000\,\mathrm{GV}$. But at both low and high R_{cr}, e'_{cr} is low, so it is quite unclear how a damping rate similar to that proposed, which is suppressed by e'_{cr}, could dominate over the entire range of interest.

Briefly, we note that an alternative damping mechanism with $\xi_{cr} \approx 0$ that scales with e_A more steeply than NLL damping ($\xi_A > 1$) would formally admit steady-state solutions [solving problem (3), per Appendix A]. But since this would give $e_A \propto (e'_{cr})^{1/(1+\xi_A)}$, it would still suffer from the spectral shape problem (2), unless one takes $\xi_A \gg 1$, with $\xi_k \sim (1+\xi_A)/2 \gg 1$. This is a rather unusual scaling. Moreover, in this limit, the required normalization of this added $\Gamma_{new,\,damp\,\pm} \propto e_A^{\xi_A}\, k_{\parallel}^{(1+\xi_A)/2}$ term at the wavenumbers and e_A observationally required at $\sim 1\,\mathrm{GV}$ would be problematically large: $\sim 10^5$ times larger than the standard NLL damping term, despite its being nominally much higher-order in e_A/e_B. As such, we will not consider this particular class of alternative damping model further.

3.5.2 Alternative driving

Alternatively, we could invoke a different or additional driving term $S_{new,\pm}$ while keeping the known damping mechanisms. Consider an alternative source term as parametrized in Section 2.6: $S_{new,\pm} = \mathrm{d}E/\mathrm{d}\ln k_{\parallel}\,\mathrm{d}t\,\mathrm{dVolume} \propto k_{\parallel}^{\zeta_k}\, e_A^{\zeta_A}$ (i.e. $\zeta_{cr} = 0$, so we avoid the problems of SC models above). Given some damping rates parametrized in similar fashion as in Section 2.5, a desired δ_s is obtained for $\zeta_k = \xi_k - (1-\delta_s)\,(1+\xi_A-\zeta_A)$.

(1) Extrinsic/External sources: First, consider the case with $\zeta_A = 0$, i.e. $S_{new,\pm} = S_{new,\,ext}$ independent of e_A, as appropriate for a truly 'extrinsic' or 'external' energy driving/pumping term (akin to ET in this limited sense). Then, if we consider ξ_A, ξ_k for all possible damping mechanisms in Section 2.5, and allow $0.3 \lesssim \delta \lesssim 0.7$, then the range of possible ζ_k that produces the desired δ_s is bounded by $-0.7 \lesssim \zeta_k \lesssim 0.5$. But more realistically, if we restrict to $\delta_s \approx 1/2$, and ignore ion-neutral damping (which has a rather different scaling ξ_k from other rates, and is rarely relevant in the volume-filling ISM which dominates the statistics as seen in the LISM), then allowing for all other damping processes requires a much narrower range of $-0.1 \lesssim \zeta_k \lesssim 0.25$. In other words, a model with $S_{new,\,ext} \sim$ constant, or weakly dependent on $k_{\parallel}$ (hence $r_{g,\,cr}$), is potentially viable. One example, which has approximately the correct normalization if we assume turbulent, linear-Landau, or dust damping dominates $\Gamma_{\pm}$, would be $S_{new,\,ext} \sim 0.01\,(v_{A,\,ideal}/\ell_A)\,e_B$ (this could be multiplied by a weak power of $k_{\parallel}\,r_{g,\,cr}$ or $r_{g,\,cr}$, e.g. $(k_{\parallel}\,r_{g,\,cr}[1\,\mathrm{GV}])^{\zeta_k}$ with $-0.1 \lesssim \zeta_k \lesssim 0.25$).

(2) Linear sources: Secondly, consider the case with $\zeta_A = 1$, i.e. $S_{new,\pm} = S_{new,\,lin} \propto e_A$. This would be appropriate for e.g. any linear instability that amplifies e_A. The gyroresonant CR instabilities (or non-resonant CR instabilities; Bell 2004) invoked in SC models are one such instability, but suffer from the problems described in Section 3.4. For any $\zeta_A \approx 1$, again allowing for a broad range of δ_s and all possible damping mechanisms bounds $0 \lesssim \zeta_k \lesssim 0.85$, but restricting to $\delta_s \approx 1/2$ and ignoring ion-neutral damping gives the much narrower range $0.4 \lesssim \zeta_k \lesssim 0.75$. Again, an example here that has approximately the correct normalization would be e.g. $S_{new,\,lin} \sim 0.001\,(v_{A,\,ideal}/\ell_A)\,(k_{\parallel}\,\ell_A)^{1/2}\,e_A$.

Either of these novel source terms seem to be at least plausible. For the first 'extrinsic source' case ($\zeta_A = 0$), the scaling $S_{new,\,ext} \sim$ constant is just that assumed in isotropic, undamped turbulent cascade models. But, the more general condition is simply that the driving in energy space – i.e. $\mathrm{d}|\delta\mathbf{B}^2(k_{\parallel})|/\mathrm{d}\ln k_{\parallel}\,\mathrm{d}t\,\mathrm{dVolume}$ – is only weakly dependent of $k_{\parallel}$ (i.e. the driving rate is comparable across scales from $\sim$MeV-TeV gyroradii). We emphasize that from the constraints in Section 2.5 and assumed structure of the competition with scale-dependent damping, this cannot apply to any of the standard physically motivated models for a traditional MHD turbulent cascade from larger scales, which would introduce the damping and anisotropy problems. Rather, in order to satisfy the requirements, it is more natural to consider modes as driven and damped effectively 'independently' on each scale, in a manner where the energy driving/injection rate is comparable per logarithmic

interval in parallel $k_\parallel$, but allowing for damping and/or anisotropy and/or transfer so long as this condition is met. Note that the required normalization/total energy driving rate in this scenario is quite small – only ~ 1 per cent of the driving/dissipation rate of ISM turbulence on larger scales.

For the second 'linear source' case ($\zeta_A = 1$), the linear scaling $S_{\rm new,\,lin} \sim \Psi_{\rm new,\,lin}\, e_A$ with $\Psi_{\rm new,\,lin} \propto k_\parallel^{0.4-0.75}$ is physically easy to imagine. Most obviously, a huge variety of multifluid instabilities present in the ISM exhibit behaviour that could lead to similar scalings. For example, at a two-fluid interface, the Rayleigh–Taylor instability (RTI) with $\Psi_{\rm new,\,lin} \sim \sqrt{g\,k}$ where g is some acceleration, would require only very weak $g \sim 10^{-6}\, v_A^2/\ell_A$ to produce roughly the correct behaviour. More generally, any 'co-spatial fluids' – i.e. any two fluids that both share the same volume, such as dust and gas, ions and neutrals, radiation and gas, etc. – can be unstable to the family of resonant-drag instabilities (RDIs; Squire & Hopkins 2018), many of which drive modes that could scatter CRs with the desired scalings.[12] For example, the Alfvén-wave dust-gas RDI (Hopkins & Squire 2018) is unstable on all scales of interest here, and has a growth rate $\Psi_{\rm new,\,lin} \sim (f_{\rm dg}\,\Delta a_{\rm dg}\,k_\parallel)^{1/2}$ at intermediate $k_\parallel$ and $\propto k_\parallel^{1/3}$ at high $k_\parallel$, where $f_{\rm dg}$ is the dust-to-gas ratio [$f_{\rm dg} \lesssim 0.01$ in the (MW), depending on gas phase], $\Delta a_{\rm dg}$ is any external acceleration felt differently by dust and gas (e.g. radiation pressure), and we have assumed the dust-gas drift speed is sub-Alfvénic. So such a mechanism would require $f_{\rm dg}\,\Delta a_{\rm dg} \sim 10^{-6}\, v_A^2/\ell_A$ in order to scatter CRs sufficiently, well within plausible ranges (see e.g. Weingartner & Draine 2001). Alternatively, very similar RDIs can arise between the ionized and neutral gas phases in partially ionized gas, several of which are studied in Tytarenko, Williams & Falle (2002) with growth rates $\Psi_{\rm new,\,lin} \propto k_\parallel^{1/3-2/3}$ depending on $k_\parallel$ (although some of these are stabilized on small scales by pressure effects in the neutrals).

These alternative source models appear more well-motivated than the alternative damping model in Section 3.5.1. To start, there are physically motivated, known processes that could potentially produce the correct additional source terms. But also, they do not require that we 'discard' major known damping mechanisms or other known source terms, in order to make these models 'work' (whereas with the alternative damping model, we must invoke some other physics to explain why other other known damping processes do not dominate). In fact, we can simply 'add' similar driving terms on top of the known terms in Section 2.5–2.7. There is, however, one remaining caveat if we do so. We still need to avoid the SC bottleneck/runaway (if we still include $S_{\rm sc}$) in regions that *do* have high $e'_{\rm cr}$, specifically at rigidities $\sim 1-10\,$GV where $e'_{\rm cr}$ is maximized. In other words, $S_{\rm sc}$ would be expected to still be large and potentially dominate over $S_{\rm new}$ by a factor as large as $\sim$100, given the 'normalization' problem in Section 2.6 (which imposing new driving terms does not solve). This is not a serious problem outside of the range $\sim 1-10\,$GV, because $e'_{\rm cr}$ is smaller. But at $\sim 1-10\,$GV, all of the plausible solutions to the 'normalization' problem discussed in Section 2.6 – e.g. more accurately calculating gyroresonant growth rates that account for the full CR spectrum, and more accurately calculating SC-induced scattering rates ($\hat{\nu}_{\rm s}$ arising from the SC) – act to reduce the SC contribution, so can potentially alleviate this problem.

[12]More technically, if a second fluid (e.g. dust, radiation, or neutrals) can 'resonate' with Alfvén waves with the desired $k_\parallel$ (i.e. if they have a natural mode with a frequency that matches that of the Alfvén wave), and any coupling that depends on their relative streaming velocities (e.g. collisions, drag, Lorentz forces), then it can produce an RDI. These generally drive modes in $k_\parallel$ that could scatter CRs and have growth rates $\gamma \sim k_\parallel^\alpha$, where $1/3 \lesssim \alpha \lesssim 2/3$ depending on the scale and type of mode (see Squire & Hopkins 2018). For the 'intermediate' $k_\parallel$ or small-$f_{\rm dg}$ case (where $f_{\rm dg}$ is the mass-density ratio of the two fluids), assuming the streaming/drift velocity is sub-Alfvénic, this generically gives $\Psi_{\rm new,\,lin} \sim (f_{\rm dg}\,\Delta a_{\rm dg}\,k_\parallel)^{1/2}$ where $\Delta a_{\rm dg}$ is any difference in the accelerations felt by the two fluids that gives rise to non-zero streaming.

4 NUMERICAL SIMULATION METHODS

We now test these analytic conclusions in detailed numerical simulations, beginning by describing our simulation methods. Briefly though, we note that there are several motivations to explore fully dynamical simulations of global galaxy formation and structure. One is to test whether these conclusions are robust in a more realistic, turbulent, multiphase medium (in which plasma properties such as v_A, etc. vary on scales which are resolved but still small compared to the size of the entire Galaxy; see discussion in Paper I) as is present in these numerical simulations, but cannot be captured in even state-of-the art analytic Galactic structure models (compare e.g. Benincasa et al. 2020; Maurin 2020). Another is to test whether non-equilibrium CR dynamics (i.e. dynamical behaviours whether either the background plasma, or CR flux or energy equations themselves, are not in equilibrium), neglected in any 'steady-state' models, could impact these conclusions (see e.g. Bustard & Zweibel 2021; Hopkins et al. 2021d; Thomas, Pfrommer & Pakmor 2022). Yet another motivation is to test whether 'feedback' or back-reaction effects from CRs on the medium, necessarily ignored in any post-processing models where CRs are not evolved 'on the fly' could somehow produce different conclusions. Examples of this include the effect of global CR pressure gradients and CR coupling to magnetic fields re-accelerating outflows to large CGM radii (Salem & Bryan 2014; Simpson et al. 2016; Wiener et al. 2017; Hopkins et al. 2021b); or producing strong shocks, mixing via bouyancy effects, or thermally heating the CGM (Enßlin et al. 2011; Wiener, Oh & Guo 2013a; Su et al. 2020, 2021; Wellons et al. 2022); or altering the phase structure of the CGM via allowing gas to occupy states prohibited in strict thermal pressure equilibrium (Salem, Bryan & Corlies 2016; Butsky & Quinn 2018; Butsky et al. 2020; Ji et al. 2020, 2021); or altering the vertical support, hence pressure balance, turbulent strength, or magnetic field strengths of galactic discs (Wiener, Zweibel & Oh 2013b; Hopkins et al. 2020b; Chan et al. 2021; Ponnada et al. 2022); or altering ionization balance in neutral gas in e.g. GMCs or galactic nuclei (Gaches & Offner 2018; Hopkins et al. 2021a; Armillotta, Ostriker & Jiang 2021). We emphasize that all of these effects have been studied in previous simulations with the same physics, code/numerical methods, and resolution (references given), demonstrating that they can in fact be captured – the difference is that these previous studies have generally treated CRs in the 'single-bin' approximation (integrating only a total CR energy, rather than the full spectrum, and neglecting differences between species) with a simple phenomenological transport/scattering rate model (e.g. a single universally constant scattering rate or diffusivity or streaming speed).[13]

[13]Briefly, it is also worth noting that even studies using entirely different codes and numerical methods and ICs have largely produced similar results to the FIRE simulations referenced here, provided they include similar physics and adopt similar CR transport parameters (see e.g. the discussion in Armillotta et al. 2021). Moreover, while there is of course some resolution-dependence and there will necessarily be unresolved scales (discussed in more detail below), these appear to have little effect on the mean properties predicted in the studies above (see references therein) and to manifest weakly via effects like the failure to resolve small molecular clouds, leading to

4.1 Non-CR physics

The simulations studied here are identical to those in Hopkins et al. (2021a; hereafter Paper I), except for the expressions used for the CR scattering rates, so we only briefly summarize the methods here. The simulations are run with GIZMO[14] (Hopkins 2015), in meshless finite-mass mode, with MHD solved as in Hopkins & Raives (2016) and Hopkins (2016) with anisotropic Spitzer–Braginskii conduction/viscosity as in Hopkins (2017) and Su et al. (2017), self-gravity solved with adaptive Lagrangian force softening, and cooling, star formation, and stellar feedback following the Feedback In Realistic Environment (FIRE)-3 implementation of the FIRE physics (Hopkins et al. 2018b, 2022a). We explicitly follow enrichment, dynamics, and chemistry of 11 species (Colbrook et al. 2017; Escala et al. 2018), cooling and non-equilibrium ionization/atomic/molecular chemistry from $\sim$1–10^{10} K, including metal-line, molecular, fine-structure, photo-electric, ionization, and other processes with local sources and a metagalactic (self-shielded) background from Faucher-Giguère (2020). Locally self-gravitating Jeans-unstable gas in converging flows is allowed to form stars following Hopkins, Narayanan & Murray (2013) and Grudić et al. (2018), and once formed stars evolve according to explicit stellar evolution models and return metals, mass, momentum, and energy to the ISM via resolved individual SNe (both Ia and core-collapse) and O/B and AGB mass-loss as in Hopkins et al. (2018a), with radiative heating and momentum fluxes solved using a five-band radiation-hydrodynamic approximation from Hopkins et al. (2020a). We note resolution tests below but the default mass resolution is $\Delta m_i \approx 7000\,M_\odot$, so the spatial/force resolution scales with density as $\Delta x_i \sim 10\,{\rm pc}\,(n/100\,{\rm cm^{-3}})^{-1/3}$, and the simulations naturally feature a multiphase ISM with hot phases at $n \ll 0.01\,{\rm cm^{-3}}$ and molecular clouds (with the mass spectrum and other scalings of the most massive, resolved clouds agreeing well with observations; Benincasa et al. 2020; Guszejnov et al. 2020; Keating et al. 2020) up to the maximum densities where the fragmentation scale can be resolved of $n \sim 10^3 - 10^4\,{\rm cm^{-3}}$ (see Hopkins et al. 2018b, for more details). The simulations here are 'controlled restarts' where we take a fully cosmological simulation run from $z \sim 100$ to $z = 0$ with a simpler CR treatment from Hopkins et al. (2020b), selected because it forms a galaxy similar in all obvious relevant properties to the MW, and restart it from a snapshot at $z \approx 0.05$, modifying the CR assumptions, and running for ≈ 500 Myr to $z = 0$. This is sufficient for all CR quantities in the ISM to reach their new quasi-steady-state values, but ensures (unlike running an entirely new cosmological simulation) that our comparison is 'controlled' (bulk Galaxy properties are similar). All numerical details of the methods are described and tested extensively in Hopkins et al. (2021a).

4.2 CR physics

Following Hopkins et al. (2022b), we explicitly evolve the CR DF $f_{\rm cr} = f_{\rm cr}(\mathbf{x},\,\mathbf{p}_{\rm cr},\,t,\,s,\,\ldots)$, assuming a gyrotropic DF following equations (1–2). By definition, $\langle\mu\rangle \equiv \bar{f}_{\rm cr,\,1}/\bar{f}_{\rm cr,\,0}$, and the moments hierarchy for $\bar{f}_{\rm cr,\,2}$ is closed by the interpolated M1-like relation $\langle\mu^2\rangle \approx (3+4\,\langle\mu\rangle^2)/(5+2\,\sqrt{4-3\,\langle\mu\rangle^2})$, which captures the exact behaviour in both the 'free streaming' or weak-scattering and isotropic-DF or strong-scattering limits (Hopkins et al. 2022b). All the variables above are functions of position and time. CRs act on the gas and radiation fields: the appropriate collisional/radiative terms are either thermalized or added to the total radiation (e.g. Bremstrahhlung, inverse Compton, etc.) or magnetic energy, and the CRs exert forces on the gas in the form of the Lorentz force (proportional to the perpendicular CR pressure gradient) and parallel force from scattering as detailed in Hopkins et al. (2022b).

The momentum-space evolution of $\bar{f}_{\rm cr,\,0}$ is integrated independently in every resolution element using the finite-momentum-space-volume scheme in Girichidis et al. (2020), treating $\bar{f}_{\rm cr,\,0}(p_{\rm cr})$ as a series of independent piecewise power-laws with exactly computed number and energy fluxes (so the scheme exactly conserves CR number and energy). We discretize with $\sim$11 independent power-law intervals (each with an evolving slope and normalization) for $\bar{f}_{\rm cr,\,0}(p_{\rm cr})$ spanning $\sim$MeV-TeV energies, per CR species, per simulation cell. We cannot resolve first-order Fermi acceleration so we model injection by assuming ~ 10 per cent of the initial pre-shock kinetic energy goes into CRs, with ~ 2 per cent of that into leptons, at the formation of the reverse shock around each SNe and/or O/B winds, with the relative number per species set by the evolved abundances at that point (e.g. the test-particle limit) with a fixed injection spectrum $j_{\rm cr}(R_{\rm cr}) \propto R_{\rm cr}^{-4.2}$. We explicitly follow the CR species protons p (H), CNO, ^{7}Be, ^{9}Be, ^{10}Be, anti-protons $\bar{p}$, electrons e^-, and positrons e^+. In the loss terms $\mathcal{R}_{\rm loss}$ and $j_{\rm cr}$, we include Coulomb & ionization, Bremstrahhlung, inverse Compton, synchrotron, pionic, fragmentation, radioactive decay, and annihilation processes, with standard cross-sections compiled in Paper I. This includes the secondary production of e.g. e^+, e^-, B, C, Be, etc. All ISM quantities needed for these rates (e.g. gas densities & ionization states, magnetic & radiation energy densities, etc.) are taken directly from the dynamically evolved simulation quantities in the cell. As also noted in Hopkins et al. (2022b), our equation (1–2) automatically include 'adiabatic' ($\mathbb{D}_{\rm cr} : \nabla\mathbf{u}_{\rm gas}$), 'streaming loss' or 'gyroresonant' ($\propto \bar{v}_A$ or $D_{\mu p}$, $D_{p\mu}$), and 'diffusive' ($\propto D_{pp}$) re-acceleration terms, in more general and accurate forms than usually considered.

We calculate $\bar{v}_{\rm s,\pm}$ following equation (6), with $\hat{v}_{\rm s} \approx 3/4$ appropriate for grey scattering, and $e_\pm$ determined from equation (7), for a given set of sources $S_\pm = \sum_i S_{i,\pm}$ and damping terms $Q_\pm = \sum_i \Gamma_{i,\pm}\,e_\pm$. We have considered both the cases where we explicitly dynamically evolve the time-dependence of equation (6) alongside the CR flux and energy equations, or where we simply set $e_\pm$ to the local-steady-state values (setting $D_t e_\pm \pm \nabla\cdot(v_A\,e_\pm\,\hat{\mathbf{b}}) \rightarrow 0$); these give very similar results for our study below, so we default to the local-state-state values as it involves slightly reduced computational expense, which allows a larger parameter survey. The key physics we vary in our tests is the scaling of the sources and damping rates $S_{i,\pm}$ and $\Gamma_{i,\pm}$, or equivalently scattering rates $\bar{\nu}_{\rm s}$.

4.3 Reference model and quantities measured

We stress from the above that (1) all of the CR physics needed to resolve, in principle, any of the known relevant CR-gas interactions or feedback effects on 'macroscopic' (simulation-resolved) scales are included; (2) all of the plasma properties (e.g. $\mathbf{B}$, n, ℓ_A, $e_{\rm cr}$) needed to calculate the 'microscopic' (unresolved, gyroscale) scattering rates in the (extrapolative) models we will consider are self-consistently predicted on the resolved simulation scales; and (3) given those (assumed) scattering rates, our simulations naturally produce a self-consistent prediction for the CR spectra across the range of energies and species we consider.

slightly more-clustered star formation and higher variability (see Hopkins et al. 2018b, 2022a; Armillotta et al. 2021), which only strengthens our ultimate conclusions since it means the simulations sample a broader range of possibilities.

[14] A public version of GIZMO is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html

The key physics of CR transport in our model therefore reduces to our expressions for the source $S_{\pm}$ and damping $Q_{\pm}$ terms in equation (7). We will explore many model variations, but it is useful to first define a 'reference model,' which attempts to represent the best current theoretical understanding of SC + ET effects as developed in e.g. Zweibel (2013), Ruszkowski, Yang & Zweibel (2017), Zweibel (2017), and Thomas & Pfrommer (2019) and other references in Section 2.5–2.7. In this 'baseline' model, we take: $S_{\pm} = S_{\rm sc,\pm} + S_{\rm et,\pm}$ where $S_{\rm sc,\pm}$ follows equation (9), and $S_{\rm et,\pm} = \alpha_t(k_{\parallel})\, e_{\rm B}\, \Gamma_{\rm turb}\, (k_{\parallel}\, \ell_A)^{-1}$ assumes an anisotropic GS95-like Alfvénic cascade (Chandran 2000). We take $Q_{\pm} = (\Gamma_{\rm in} + \Gamma_{\rm dust} + \Gamma_{\rm turb/LL})\, e_{\pm} + \Gamma^{0}_{\rm nll}\, (e_{\pm}/e_{\rm B})\, e_{\pm}$ where the expressions for $\Gamma_{\rm in}$, $\Gamma_{\rm dust}$, and $\Gamma^{0}_{\rm nll}$ are in Section 2.5, and for consistency with our driving terms $S_{\pm}$ (since we are assuming parallel modes being sheared out by a GS95-type cascade), we have $\Gamma_{\rm turb/LL} = [(v_{A,\,\rm ideal} + 0.4\, c_s)/\ell_A]\, (k_{\parallel}\, \ell_A)^{1/2}$ (where the $0.4\, c_s$ term is the 'linear Landau' term). We use the appropriate $v_{A,\,\rm eff}$ in equation (5) for the relevant v_A terms in the CR equations, and note that these expressions self-consistently determine the relation between $\bar{v}_A$ and v_A.

We will focus on comparison of the models here to the Solar-neighborhood/LISM constraints – the only place where the full CR spectrum of various species can be determined. In Paper I (where we study only phenomenological CR transport models), we consider a more extensive suite of constraints, including spatially resolved γ-ray emission and ionization constraints that span various Galactic environments, as well as comparisons of different CR species and abundances not shown here. While of course any ultimate 'successful' model must produce agreement with *all* of these constraints, our focus here is ruling out a number of models that *cannot* reproduce the observations, for which a simpler comparison of the LISM spectral shapes and secondary-to-primary ratios is both sufficient and most useful, given that the theoretical slope δ_s most directly manifests in the shape of the predicted secondary-to-primary ratio as a function of energy. The details of how we compare to observations are given in Paper I, but briefly we select all gas cells in a mock Solar circle (at galactocentric radii $r = 7 - 9\,$kpc), in the mid-plane ($|z| < 1\,$kpc), with gas densities similar to those observed ($n \sim 0.3 - 3\,{\rm cm^{-3}}$), and calculate the median CR spectrum of all gas in this ensemble. To define the 'scatter', we allow for a wider range of both galactocentric radii ($4 - 10\,$kpc) – allowing for the fact that our galaxies are not perfect MW analogs – and a wider range of densities ($n = 0.1 - 10\,{\rm cm^{-3}}$) and compute the interquartile ranges of all CR spectra in all cells meeting these criteria. Of course, we expect CR spectra to vary with Galactic environment, and this is discussed extensively in Paper I. We further have examined all of our predicted CR spectra in both different Galactic annuli from $r = 1 - 15\,$kpc and at different densities $0.001 - 10\,{\rm cm^{-3}}$, as well as gas selected only in different thermal phases (though this is closely related to density selection as shown in Paper I); importantly, while the normalization and detailed spectral shape of the CRs can depend on these environmental properties, none of our conclusions (particularly about the shape of B/C and δ_s, and the success or failure of different models) depends on exactly where or how we measure the CR spectra.

4.4 Model variations considered

We have tested a large number of model variations in our simulations (many in concert with one another), in order to systematically survey whether different changes to our default model could resolve the qualitative tensions described above. Here, we outline variations considered, grouping them into those that have no appreciable effect on the *qualitative* behaviours of interest in this paper, and those that we find to be most significant.

4.4.1 *Variations that do not alter our qualitative conclusions*

The following variations – all of which we have tested in full simulations to verify the robustness of our results – do not alter our qualitative conclusions, even if they produce systematic or quantitative shifts in predicted quantities. We therefore will not discuss them in detail below, but list them for completeness here.

(i) Changing Galaxy and Stellar Assumptions: as studied in detail in Paper I for simpler power-law scattering rates, we have rerun adopting two different fiducial MW-like galaxy simulations as our IC (**m12f** and **m12m**, instead of our usual default **m12i** here), all of which are similar to the real MW but differ in various details (Garrison-Kimmel et al. 2019; Samuel et al. 2020). Also, as in Paper I, have also arbitrarily multiplied the magnetic fields in our **m12i** ICs by 10 and 0.1 (even though the 'default' values agree well with MW observations; Su et al. 2018; Guszejnov et al. 2020), as these are both theoretically and observationally uncertain and influence the transport physics. Finally, we have also rerun using both our FIRE-3 (Hopkins et al. 2022a) and the older FIRE-2 (Hopkins et al. 2018b) implementation of the FIRE physics, the latter of which uses older stellar evolution and cooling tables leading to slightly different SNe and stellar mass-loss rates, detailed cooling physics, etc. As shown in Paper I, these make significantly smaller differences compared to changing CR transport coefficients at the level of detail considered here.

(ii) CR Injection Parameters: we have systematically varied the injection spectrum, e.g. considering slopes $j_{\rm cr} \propto R_{\rm cr}^{-\psi_{\rm inj}}$ within a broad range of $\psi_{\rm inj} = 3.2$–5.2, allowing for a 'broken' power-law with a break at $\sim 1\,$GV, freeing the normalization of the injected energy fraction and normalization of different components (e.g. leptonic versus hadronic). These variations are again discussed in detail in Paper I; they can be used to improve the agreement of a given model with observed CR spectral shapes, but do not resolve the qualitative problems that are evident in secondary-to-primary ratios.

(iii) Alfvén Speeds and Streaming: we have considered replacing the full expression $v_{A,\,\rm eff}$ [equation (5)] for the gyroresonant Alfvén speed with either the 'ion Alfvén speed' ($v_{A,\,\rm ion}$), which is nearly identical to $v_{A,\,\rm eff}$, or with the ideal Alfvén speed $v_{A,\,\rm ideal}$, which is much lower in dense neutral gas. While the latter has non-negligible quantitative effects (see Paper I), because the overwhelmingly neutral gas has a relatively small volume filling-fraction (so contributes only modestly in a weighted sense, for diffusive CRs), it does not alter the qualitative conclusions here regarding the success or failure modes of different models. We have also varied the 'streaming speed' $\bar{v}_A \equiv v_A\, (\bar{v}_{\rm s,+} - \bar{v}_{\rm s,-})/(\bar{v}_{\rm s,+} + \bar{v}_{\rm s,-})$, which we by default solve for explicitly, by replacing it with either exactly $|\bar{v}_A| = 0$ (the expectation in e.g. 'pure ET models') or $|\bar{v}_A| = v_{A,\,\rm eff}$ (the expectation in 'pure SC models'). This again has little effect, as this is generally subdominant to the diffusive or 'super-Alfvénic' streaming speed (Evoli et al. 2017; Amato & Blasi 2018; Chan et al. 2019; Su et al. 2020; De La Torre Luque et al. 2021; Hopkins et al. 2021d).

(iv) Renormalizing or Disabling Different 'Reference Model' Terms: we have rerun our reference model, multiplying each source term ($S_{\rm sc}$, $S_{\rm et}$) and damping term ($\Gamma_{\rm in}$, $\Gamma_{\rm dust}$, $\Gamma_{\rm turb/LL}$, $\Gamma_{\rm nll}$) by 100 and by 0.01 or 10^{-10} (effectively disabling it entirely). While

this can have large effects in some cases (discussed below), and ameliorate some of the 'normalization' tensions described above, none of these modifications, in and of themselves, resolves the fundamental issues of the failure of SC or ET models (whichever is dominant) – i.e. there is no 'single term' which drives the qualitative problems discussed above, and only the 'normalization' problem is substantively addressed by these renomalization experiments (see also Hopkins et al. 2021d).

(v) Variant ET Models: we have experimented with a number of variant 'pure ET' models (disabling the SC source term $S_{\rm sc}$) or ET + SC models (retaining $S_{\rm sc}$) that vary $S_{\rm et,\pm}$ (and $\Gamma_{\rm turb}$ which must match appropriately). For each, we have considered both (a) retaining all the other, usual damping $\Gamma_{\pm}$ terms (e.g. $\Gamma_{\rm in}$) or (b) disabling all damping terms other than the cascade transfer term $\Gamma_{\rm turb}$, so the spectrum is exactly that predicted by classical ET models. We consider each of the ET models reviewed in Hopkins et al. (2021d) and Section 2.7: (1) 'standard' Alfvénic turbulence (our default $S_{\rm et,\pm}$); (2) 'Alfvén-Max,' which assumes an anisotropic Alfvénic cascade but arbitrarily sets $\alpha_t = 1$ (this ignores the gyro-averaging correction for anisotropic modes, but still retains the effect of anisotropy producing an $e_A \propto k_{\parallel}^{-1}$ spectrum); (3) 'YL04' which follows Yan & Lazarian (2004) as detailed in appendices of Hopkins et al. (2021d), accounting for collisionless and viscous damping, and accounting for the much stronger effects of damping (super-exponentially suppressing $S_{\rm et,\pm}$) when the neutral fraction is non-zero or plasma $\beta_{\rm plasma} \equiv c_s^2/v_{A,\,\rm ideal}^2 > 1$, as well as a variant where we neglect the predicted suppression from ion-neutral damping or plasma $\beta_{\rm plasma} > 1$ (the 'Fast-Max' or 'YL04-Max' model from Hopkins et al. 2021d); (4) a model that assumes a critically balanced Alfvénic cascade but with a modified cascade rate (which might be motivated by alignment/intermittency effects, e.g. $\alpha = 1$ in the notation of Boldyrev 2005 or $\delta = 1/8$ in Schekochihin 2022), which can modify the perpendicular spectrum significantly, but again (necessarily) gives an $e_A \propto k_{\parallel}^{-1}$ parallel spectrum with only a weakly modified $\Gamma_{\rm turb} \rightarrow (v_{A,\,\rm ideal}/\ell_A)\,(k_{\parallel}\,\ell_A)^{\xi_k}$ (we take $\xi_k = 0.4$ as a somewhat ad hoc example, for the sake of comparison).

(vi) Numerical Variations: we have considered a number of numerical variations, including (1) replacing the more general second-moment closure relation from Hopkins et al. (2022b) with the assumption that the CRs are always near-isotropic (as in Thomas & Pfrommer 2019), or (2) in flux-steady state (reducing the CR equations to a diffusion + streaming equation; Zweibel 2013), (3) testing different 'reduced speed of light' numerical approximations from $\tilde{c} \sim 0.01 - 1\,c$ to ensure convergence, (4) comparing a re-simulation at 8x improved mass resolution (2x improved force resolution), or (5) directly integrating or assuming local steady-state for the scattering modes ($D_t e_{\pm} \rightarrow 0$, as discussed above). As shown in Paper I, and Chan et al. (2019) in more detail, these have quite weak effects on our results.

4.4.2 *Variations that matter*

The variations we have studied that do lead to important results, discussed below, are summarized here.

(i) 'High' or 'Low' Initial CR Energies: As described above, in our 'default' ICs, we initialize the total CR energy density following the consistent result of a cosmological simulation, with spectral shapes matched to those observed, but these quickly converge to new equilibria. However, we have also experimented with a 'low start' and 'high start' IC, in which we multiply the initial CR energy (renormalizing all spectra) by factors of 0.001 and 10, respectively. This does not change our conclusions and neither IC resolves the SC or ET problems above; and for models where the source term $S_{\pm}$ does not depend on CR energy (non-SC-dominated), this has little effect (the simulations converge to the same equilibrium, independent of this choice). However, for SC models, where $S_{\rm sc,\pm} \propto e'_{\rm cr}$, we will show that this determines which 'attractor' solution the SC model converges towards, as described in the 'instability' problem (Section 3.4) for SC.

(ii) Adding New Damping Mechanisms: we experiment with several variant models where we add a new damping term $\Gamma_{\rm new,\,damp,\,\pm}$ (optionally disabling other damping terms in our 'reference' model $\Gamma_{\rm in}$, $\Gamma_{\rm dust}$, $\Gamma_{\rm turb/LL}$, $\Gamma_{\rm nll}$, in turn), motivated by the discussion in Section 3.5.1. In the most interesting of these experiments, we add a new damping term with the form $Q_{\rm new,\,\pm} = \Gamma_{\rm new,\,damp,\,\pm}\,e_{\pm}$ with $\Gamma_{\rm new,\,damp,\,\pm} = f_{\rm ISM}^{\Gamma}\,k_{\parallel}^{\xi_k}\,(e'_{\rm cr}/e_{\rm B})$, where $f_{\rm ISM}^{\Gamma}$ and $\xi_k \sim 1/2$ are varied as described below.

(iii) Adding New Sources: we experiment with variant models where we add a new source term, considering both 'external' and 'linear' sources motivated by the discussion in Section 3.5.2, with $S_{\pm} = S_{\rm new,\,\pm} = f_{\rm ISM}^{S}\,k_{\parallel}^{\zeta_k}\,e_{\pm}^{\zeta_A}$ with $\zeta_A = 0$ (external) or $\zeta_A = 1$ (linear) and $f_{\rm ISM}^{S}$, ζ_k varied as described below. We again consider both this added directly on top of our reference model, or disabling/renormalizing various others of the 'reference' source or damping terms in turn.

5 RESULTS: MODEL COMPARISON

We now examine the results of the full simulations. To remind the reader, these self-consistently follow the dynamics of cosmological magnetized gas inflow into galaxies, cooling, and star formation, followed by stellar evolution, stellar mass-loss in O/B and AGB winds, radiative heating and photo-ionization, and, for massive stars, supernova explosions, which inject a spectrum of CRs back into the ISM alongside energy and momentum that drive galactic outflows. Phenomena such as galactic winds, turbulence, clumping, magnetic dynamo amplification, and the like are followed self-consistently. In this medium, the injected CRs propagate according to the full dynamics equations [e.g. equations (1–2), incorporating diffusion, streaming, adiabatic gains/losses, diffusive re-acceleration, catastrophic losses, radiative losses, and the like], producing secondary and tertiary CRs 'on the fly' while they propagate. Importantly, the CRs interact directly with the gas dynamically as they travel (via momentum exchange and scattering and heating and ionization), which allows not only for the non-linear development of coupled CR-gas instabilities, but also CR-driven winds and outflows, CR heating altering star formation or ionization coupling to thermochemistry, and other unique phenomenology. The CR scattering rate $\bar{\nu}_{\rm s}$ for each rigidity is calculated self-consistently at every distinct position and time, according to the different models we explore (based on the local plasma properties). We do not enforce or assume any 'steady-state' assumptions, so non-equilbrium and non-linear phenomena can occur. We wish to understand whether this could change our key conclusions above.

With these simulations, we specifically consider the most relevant model variations to test the analytic predictions developed above. First, for reference, Fig. 1 shows an example of an empirically calibrated model (as Section 3.1) where one does *not* solve for the CR scattering rate $\bar{\nu}_{\rm s}$ according to any physical model at each position and time, but simply assumes or imposes a phenomenological uniform-in-time-and-space power-law scattering rate, $\bar{\nu}_{\rm s} \sim 10^{-9}\,{\rm s^{-1}}\,\beta_{\rm cr}\,(R_{\rm cr}/{\rm GV})^{-0.6}$, i.e. $\delta_{\rm s} = 0.6$. This was studied in Paper I,

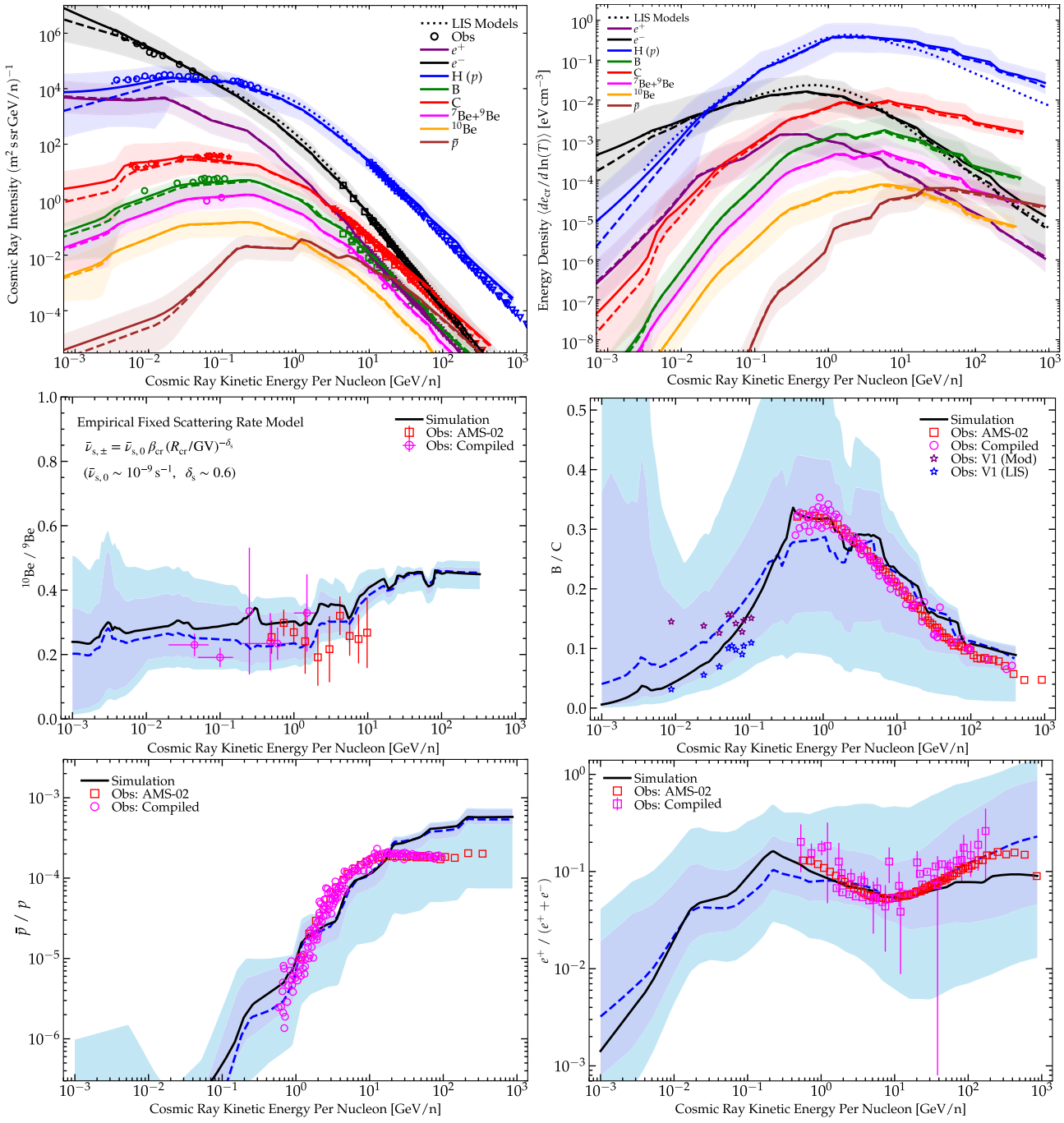

Figure 1. Example of an 'empirical model' from Paper I (see Section 3.1) where CR scattering rates are assumed to be a simple constant power-law function of rigidity $\bar{\nu}_s = 10^{-9}\ \mathrm{s}^{-1}\ \beta_{cr}\ R_{GV}^{-0.6}$. CR spectra are calculated by integrating CR dynamics and losses in a full live galaxy-formation simulation at redshift $z = 0$ (Section 4). *Top:* CR intensity/kinetic energy density spectra for different species (labelled). Lines show median (*dashed*) and mean (*solid*) values in the simulation for LISM gas in the Solar circle ($r = 7 - 9$ kpc) with density $n = 0.3 - 3\,\mathrm{cm}^{-3}$. Shaded dark (light) range corresponds to $\pm1\sigma$ ($\pm2\sigma$) range, allowing for a broader range of galacto-centric radii ($4 - 10$ kpc) and LISM densities ($n = 0.1 - 10\,\mathrm{cm}^{-3}$). Points show compiled observations (see text; Section 5). *Middle Left:* ^{10}Be/^{9}Be; dark purple (light cyan) range shows the $\pm1\,\sigma$ ($\pm2\,\sigma$) range. *Middle Right:* B/C ratio. *Bottom Left:* $\bar{p}/p$ ratio. Note the value at the highest-energies is significantly affected by our upper boundary (we do not evolve p or heavier ions with rigidity $\gtrsim 1000$ GV). *Bottom Right:* $e^+/(e^+ + e^-)$ ratio. All of these properties can be reasonably well-fit with a simple empirical power-law scaling. The spectra are nearly independent of the initial CR spectra after $\sim$100 Myr of evolution.

where we show explicitly that similar quality fits could be obtained for a narrow range of $\delta_s \sim 0.5 - 0.7$, independent of all the parameters listed as 'unimportant' in Section 4.4.1, as well as the normalization ('high' or 'low') of the CR energies in the ICs (i.e. the system rapidly converges to the same steady-state results, independent of the details of the IC).

We compare the predicted spectra of a variety of species including H (p), $\bar{p}$, e^+, e^-, B, ^{7}Be, ^{9}Be, ^{10}Be, C, N, O, and various secondary-

to-primary and radioactive species ratios. As discussed in Paper I, the most constraining combination of constraints comes from fitting the overall shape and normalization of the p and e^- spectra (which dominate $e'_{\rm cr}$, CR ionization, and γ-ray emission), the positron-to-electron and B/C ratio (which give standard secondary-to-primary ratios but depend differently on some model parameters owing to their different sensitivity to e.g. leptons versus hadrons and different loss processes), and ^{10}Be/^{9}Be (which as a diagnostic of radioactive species provides an independent 'clock,' as compared to the secondary ratios that are more sensitive to grammage). We compare the model to the observations compiled and discussed in Paper I (see that paper for more detailed discussion of the comparison, along with comparisons to a range of other observables including spatially resolved Galactic γ-ray and ionization constraints). In Fig. 1, points show observations (colours denote species), from the LISM Voyager (*circles*; Cummings et al. 2016), AMS-02 (*squares*; Aguilar et al. 2018, 2019a, b, and references therein), and compiled from other experiments, including PAMELA, HEAO, BESS, TRACER, CREAM, NUCLEON, CAPRICE, Fermi-LAT, CALET, HESS, DAMPE, and *ISO*MAX (*pentagons*; Engelmann et al. 1990; Shikaze et al. 2007; Boezio et al. 2000; Obermeier et al. 2011; Adriani et al. 2014; Abdollahi et al. 2017; Boezio et al. 2017; H. E. S. S. Collaboration 2017; Yoon et al. 2017; DAMPE Collaboration 2017; Adriani et al. 2018; Atkin et al. 2019). For the non-Voyager data, we omit observations at energies where the Solar modulation correction is estimated to be important (see Bindi et al. 2017; Bisschoff et al. 2019, and references therein). For the Voyager data, we show both the directly observed values and the 'modulation-corrected' value from Strong, Moskalenko & Ptuskin (2007) who consider models where modulation could still be important for V1 data (note this would also reduce the value of B/C observed at ~ 1 GeV).

More extensive comparisons to other observations, including e.g. observed γ-ray emissivities and CR ionization rates as a function of Galactic position, are presented in Paper I, all of which demonstrate consistency between this particular model and the observations. In future work, it will be important to compare some of the proposed alternative models below to this extended set of constraints as well.

5.1 Default SC & ET models: confirmation of failure modes

Having shown in Fig. 1 that it *is* possible to simultaneously reproduce the observations with a simple phenomenological model, we will now show that it is remarkably difficult to achieve the same in physically motivated SC or ET models. We compare the same observations from Fig. 1 to our 'default' model (Fig. 2), SC-dominated models starting from lower and higher CR energy densities (Fig. 3), and ET-dominated models (Fig. 4), defined as in Sections 2 and 4.3.

5.1.1 SC models

First, we can immediately confirm that in our 'reference model' from Section 4.3, the total scattering rate driving is dominated by the SC terms $S_{\rm sc,\pm}$ (as compared to the Alfvénic $S_{\rm et,\pm}$). This is expected since the theoretically favoured scattering rate from Alfvénic turbulence accounting for anisotropy (Section 2.7) is equivalent to a diffusivity $\kappa \gtrsim 2\times10^{33}\,{\rm cm^2\,s^{-1}}\,(\ell_A/100\,{\rm pc})$ independent of $R_{\rm cr}$ (and larger with damping). So our qualitative conclusions and the key results in Figs 2–3 are identical whether we consider 'SC + (favoured) ET' or 'SC only' ($S_{\rm et,\pm}\rightarrow 0$) models.

Figs 2 and 3 illustrate the fundamental 'instability' or 'solution collapse' problem of SC models, as discussed in Section 3.4 and derived more rigorously in Appendix A. For either the regular or 'low' or 'high' start ICs (Section 4.4.2), the system initially rapidly converges to the approximate 'local steady-state' scattering rates (i.e. the steady-state scattering rates assuming the CR and plasma properties are frozen at their instantaneous values), which allow for 'super-Alfvénic' streaming at some finite multiple of the Alfvén speed (see Appendix B). However, this is not a steady-state solution for the CR *energy density* equation, and the system then (on the CR transport time-scale ~ 10 Myr) collapses to one of the only two truly stable steady-state SC solutions for $e'_{\rm cr}$. If the initial CR energy density at some rigidity $e'_{\rm cr}$ is too low (and therefore also the SC-driving strength $S_{\rm sc}\propto e'_{\rm cr}$, and the resulting scattering rates $\bar{\nu}_{\rm s}$), the CRs escape more rapidly, further lowering $e'_{\rm cr}$, until the system collapses to the 'free streaming limit' with no scattering (the tiny residual scattering in Fig. 2 is driven by the small ET term). This occurs at all CR rigidities in our 'low start' (lower initial CR energy density) ICs and rigidities $\gtrsim 100$ GV (where $e'_{\rm cr}$ is still relatively low) in our 'normal start' ICs. On the other hand, if the initial CR energy $e'_{\rm cr}$ is too high, the system over-scatters ($\bar{\nu}_{\rm s}$ becomes very large) slowing transport and producing a bottleneck until the system collapses to the 'infinite scattering' limit, where CRs can only stream at the Alfvén speed. This gives a momentum-independent CR escape time of $\sim 1\,{\rm Gyr}\,(\ell_{\rm R,\,halo}/10\,{\rm kpc})\,(v_A/10\,{\rm km\,s^{-1}})^{-1}$ (where $\ell_{\rm R,\,halo}$ is the maximum of either the galacto-centric radius or height of the CR scattering halo), which is orders of magnitude larger than observationally allowed. The dependence of escape time on rigidity is also qualitatively different from that required by observations. This produces an order-of-magnitude excess, as well as the wrong shape/rigidity-dependence, in the CR spectrum and secondary-to-primary ratios.

We have also compared these models to observed Galactic γ-ray emissivities and ionization rates, following the identical procedure to Paper I (figs 11 &12 therein) where we compared the phenomenological model in Fig. 1 to data from Digel et al. (2001), Ackermann et al. (2011), Tibaldo (2014, 2015), Indriolo et al. (2015), Acero et al. (2016), Yang, Aharonian & Evoli (2016), and Tibaldo, Gaggero & Martin (2021). We do not show this explicitly as the information is redundant with that in Figs 2 and 3: the 'default' and 'high-start' models, which lead to the over-confined limit for $\sim 0.1-10$ GV protons that dominate the γ-ray emissivity observed, produce a γ-ray emissivity ($\propto e'_{\rm cr}\,\rho$) about a factor $\sim$30 larger than observed at Galactocentric radii $\sim 1-10$ kpc. Conversely, the 'low-start' model produces an emissivity a factor $\sim$100 lower than observed. Note that even if the CR proton spectrum in Figs 2 and 3 is lower in some low-density ISM gas or at larger galactocentric radii, or even if we uniformly increased the Alfvén speed of self-confined CR streaming by an arbitrary large factor $\sim$10–30, it is particularly hard to avoid severely over-predicting the γ-ray flux in the 'default' or 'high-start' models: it only requires some dense regions where the local Alfvén speed is low to produce excessive γ-ray emission (see Hopkins et al. 2021d). Although the CR ionization rates show the same qualitative trend, being over-predicted where the CR spectrum at low energies is high, they are less constraining. This is because low-energy CRs are well-confined (have slow diffusion) even in the phenomenological model in Fig. 1, and losses can regulate their residence time.

We also clearly see in Fig. 2 the 'spectral shapes' problem predicted in Section 3.4 for the 'normal start' model, where the CR proton and electron spectra are much too-sharply peaked around ~ 1 GV. In other words, the shape is 'too steep' at high energies

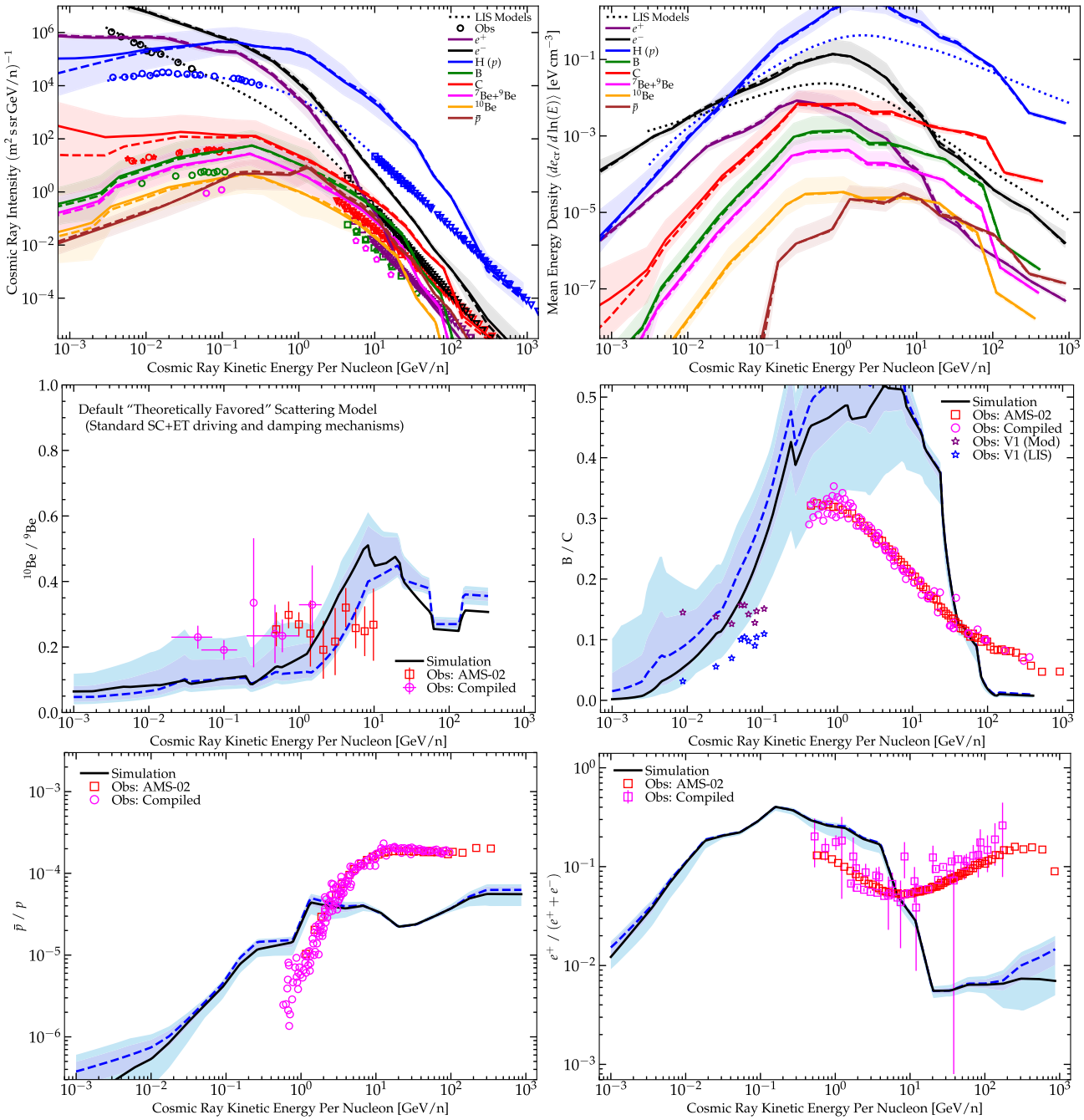

Figure 2. Predicted CR spectra as Fig. 1, but for our default, 'theoretically favoured' model. Here, the driving term $S_{\pm}$ for scattering includes both SC ($S_{sc,\pm}$) and ET ($S_{et,\pm}$) terms, accounting self-consistently for anisotropy and damping in both, but the ET terms contribute negligibly at the energies plotted (so our results are similar to a 'pure SC' model with $S_{et} = 0$). The ICs have CR spectra set to observed values (the 'normal start' in Section 4.4.2). We include all standard mode-damping mechanisms (Section 2.5). From this IC, the initially super-Alfvénic streaming at intermediate CR energies (where the CR energy density e'_{cr} is relatively high) quickly collapses to the 'bottleneck' or 'infinite scattering' solution that gives very slow CR transport (limited by the Alfvén speed, and independent of rigidity), over-predicting B/C, and e^+/e^- and e'_{cr} by an order of magnitude at $\sim 1-30$ GV. At higher CR energies ($\gtrsim 30-100$ GV) where initial e'_{cr} is lower the solutions collapse to the 'unconfined' or 'free-streaming-at-c' solution with negligible scattering (giving too-low B/C and e^+/e^-). This is the 'instability' or 'solution collapse' problem (Section 3.4): regardless of IC or renormalizing the SC driving or damping rates, no stable intermediate solutions between these extremes exist in the context of standard SC models.

and 'too shallow' at low energies. This corresponds to the effective δ_s being 'too low' at low energies and 'too high' at high energies.

We have confirmed that none of the variations in Section 4.4.1 make any appreciable difference to these behaviours. Changing, for example, the normalization of SC source or damping terms, removing damping terms, or changing the wavelength-dependence of the SC driving or damping terms, only shifts the value of e'_{cr} that divides the two unstable 'solution collapse' limits. In other words, if we systematically lower the normalization of $S_{sc,\pm}$ at

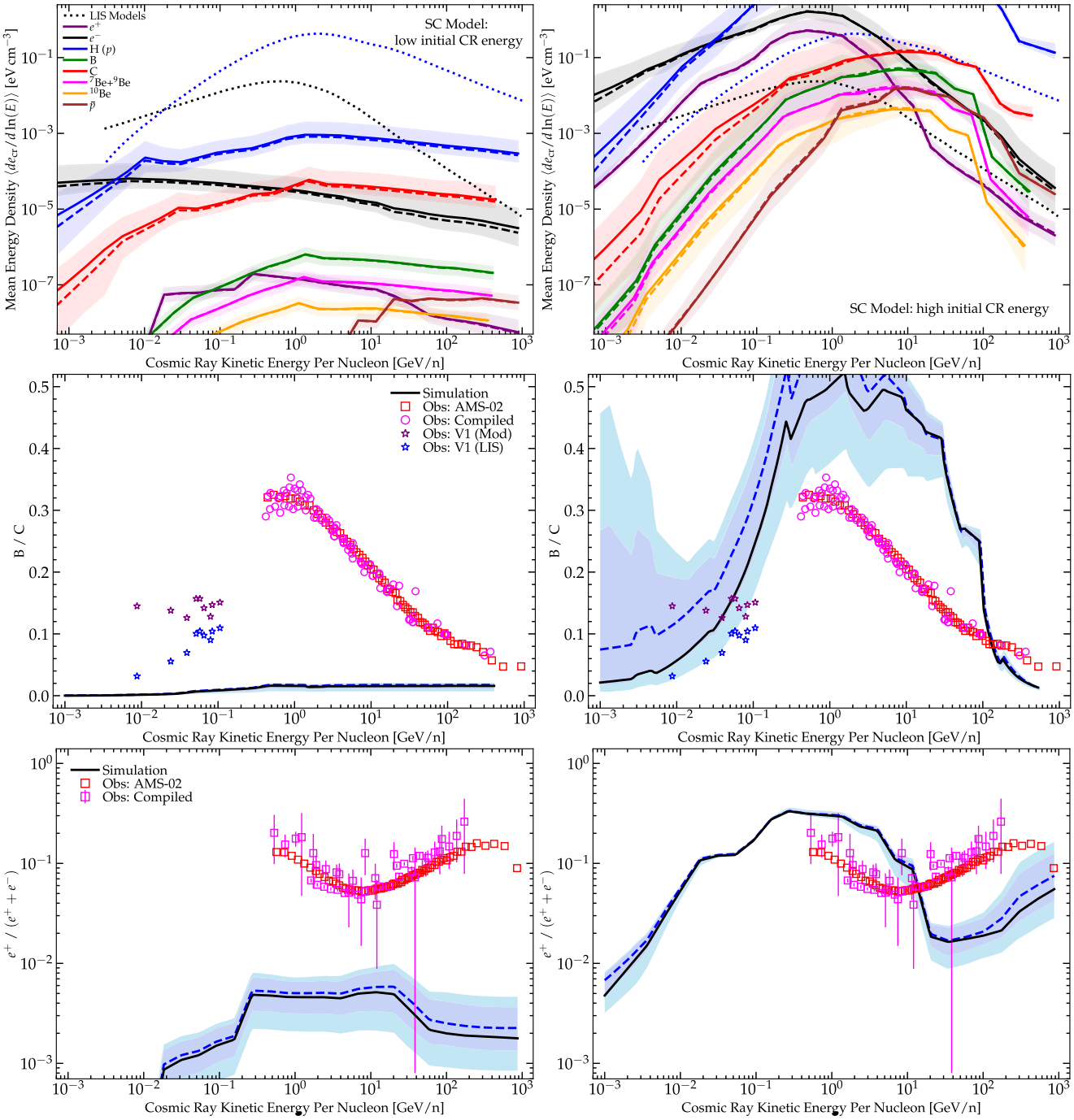

Figure 3. As Fig. 2, but for SC models that adopt a lower or higher initial CR energy (see Section 4.4.2). *Left:* a 'low-start' IC where we multiply the initial e'_{cr} by ∼0.001 relative to observed values. Now CRs at all energies collapse to the unconfined solution (the small residual scattering is from the non-zero ET terms). *Right:* a 'high-start' IC where we impose initially flat CR spectra with total CR energy multiplied by ∼10 relative to observed. Now a broader range of CR energies collapse to the 'bottleneck' solution, near the calorimetric limit, except for the highest-energy CRs that collapse to the unconfined solutions (note the high-energy leptonic spectra are strongly modified by losses here). For simplicity, we omit the plots of ^{10}Be/^{9}Be, $\bar{p}/p$, and the CR intensity: these disagree with observations in the same manner as expected from B/C, e^+/e^-, and CR spectra shown, so the information is redundant.

some wavelength $k_\parallel$ or rigidity R_{cr} by a factor A, then collapse to the 'unconfined' solution as compared to the 'infinite scattering' solution will occur at a factor ∼A lower CR energy density e'_{cr} at that R_{cr}. We also confirm that no variant model we test is somehow able to exactly balance at the 'dividing line' between the two regimes. This is not surprising: even if we could contrive a model that was balanced in this respect, our simulations are dynamical so the local CR energy density varies (as e.g. clustered SNe explode and star formation rates vary across the Galactic disc), and such a solution is unstable to variations in Galactic properties (Appendix A). So, the system is perturbed and immediately collapses into either extreme.

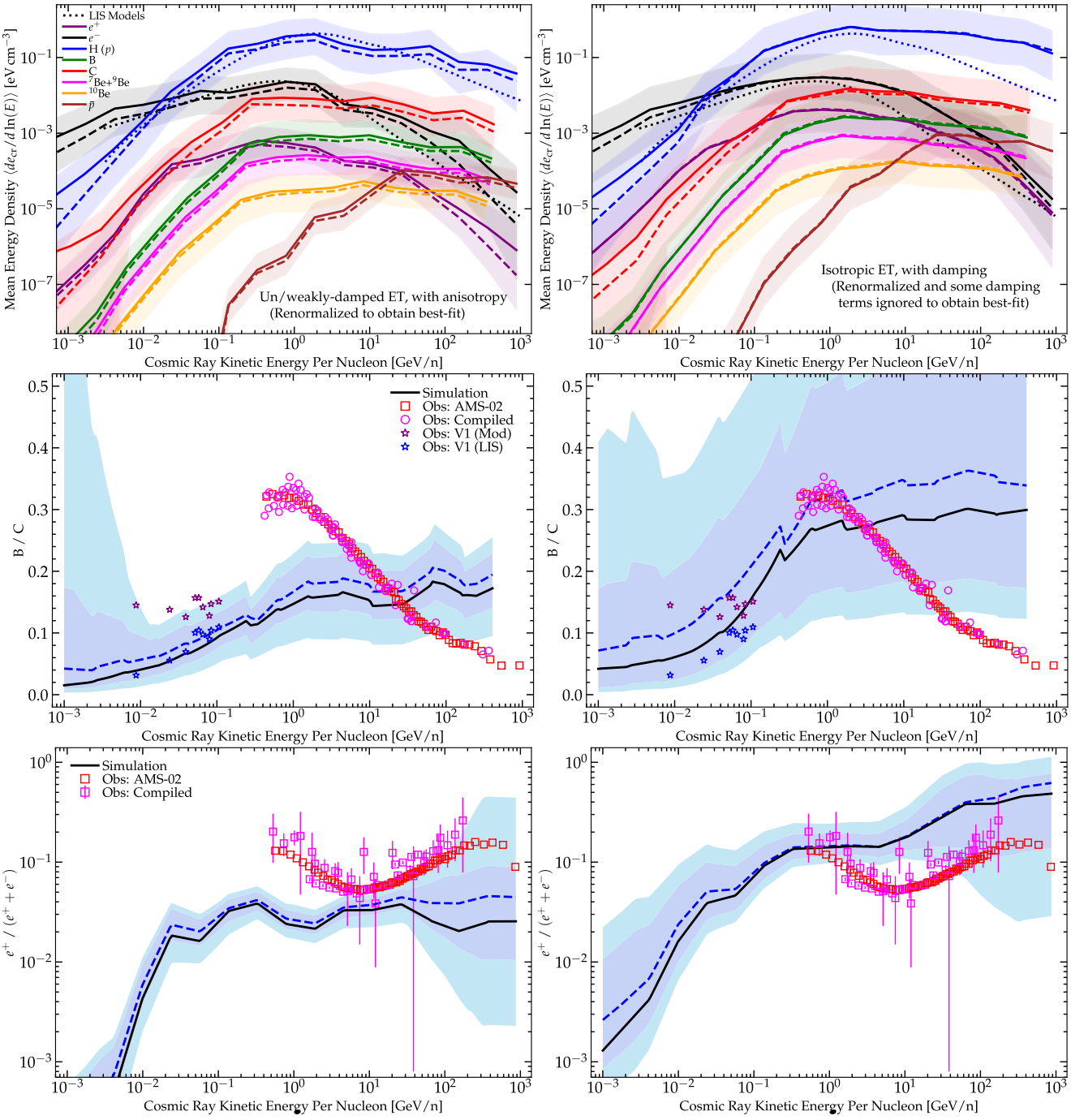

Figure 4. CR spectra (as Fig. 3) now for pure-ET models, demonstrating the anisotropy, normalization, and damping problems (Section 3.2). Here, we disable SC driving ($S_{sc,\pm} \rightarrow 0$) and arbitrarily renormalize the ET driving ($S_{et,\pm}$) or damping rates to force the models to match CR proton spectra at $\sim$1 GV. *Left:* representative behaviour of any model that accounts for anisotropy in the Alfvénic cascade (e.g. any model obeying critical balance, such as GS95). This imposes $S_{et,\pm} = \alpha_t\, e_B\, \Gamma_{turb}\, (k_\parallel\, \ell_A)^{-1}$ (Section 2.7). We renormalize $\alpha_t \sim 1$ (a factor $\sim 10^3$–10^6 larger than theoretically favoured) to fit spectra at $\sim$ 1 GV. But the spectral shapes are still incorrect. Accounting for damping or deviating from critical balance makes the disagreement worse (see Section C). *Right:* representative behaviour of any model assuming an isotropic fast-magnetosonic cascade, accounting for the fact that the spectrum is modified by damping/dissipation on a scale λ_{diss} larger than the gyro scale (here following YL04; Section 2.7). We renormalize to fit the p spectrum by disabling ion-neutral, dust and other parallel or Alfvénic damping terms and adopting the YL04 scalings for plasma $\beta_{plasma} < 1$ everywhere (otherwise the scattering rate is reduced by $\sim 10^6$). Even renormalizing to force a reasonable mean scattering rate at $\sim$GV, these models cannot reproduce observed spectral shapes: accounting for anisotropy and/or finite dissipation scales forces ET scattering rates (therefore B/C) to be *independent* or even *increasing* functions of CR energy at $\gtrsim$ GV (e.g. $\delta_s \leq 0$), contradicting observations.

For this reason, the results here are also insensitive to resolution or other micro-physical details of how we initialize the simulations (changing the magnetic field strengths, phase structure, resolution, etc.) – since there are only two limits to which the simulations can collapse (each of which contradicts observations) we can only indirectly influence 'which branch' is collapsed, or the absolute value of the Alfvén speed in the over-confined limit (which will change the exact normalization of some predictions, but not the qualitative prediction of momentum-independent escape times far in excess of those observationally allowed).[15]

Moreover, even if we take an arbitrarily renormalized SC model, and we choose to measure the CR spectra in low-density gas in the Solar circle, such that we can reproduce roughly the correct normalization of CR spectra and B/C at ~ 1 GV in the 'infinite scattering' (Alfvénic-streaming or 'high-start') limit (it is not possible to reproduce these under any circumstances in the 'free escape' limit), we (1) still see the 'spectral shape' problem and 'solution collapse' at energies far from $\sim$1 GV, (2) see solution collapse at $\ll$1 GV in different environments such as molecular clouds, which would violate observational constraints on CR ionization rates (Indriolo et al. 2009; Padovani et al. 2009; Indriolo & McCall 2012; Indriolo et al. 2015), and (3) find that for this normalization at the Solar circle, the fact that most star formation and SNe occur in the MW at radii much closer to (<5 kpc from) the Galactic centre, where gas densities are higher, leads to the prediction that the γ-ray emission at $\sim 1-10$ GeV from the Galaxy would be at nearly the proton-calorimetric limit, a factor $\sim$10–100 larger than observed in the MW and other Local Group galaxies (see discussion in Lacki et al. 2011; Blasi & Amato 2012; Evoli et al. 2017; Fu, Xia & Shen 2017; Amato & Blasi 2018; Lopez et al. 2018)

5.1.2 *ET models*

By turning off SC driving, we now examine 'pure ET' models in Fig. 4. While we have tested them to verify this, the 'most theoretically favoured' models for ET driving from either Alfvénic turbulence (realistically accounting for anisotropy following e.g. Chandran 2000; Boldyrev 2006; Lazarian 2016) or fast/magnetosonic turbulence (realistically accounting for damping following e.g. Yan & Lazarian 2002; Cho & Lazarian 2003; Yan & Lazarian 2004, 2008) are not interesting, as (at these CR energies $\sim$MeV-TeV) they predict extremely low and approximately rigidity-independent scattering rates, which correspond to diffusivities $\kappa \gtrsim 10^{33}\,{\rm cm^2\,s^{-1}}$.

[15]Briefly, we note that in future work it will be particularly interesting to explore the behaviour of recently discovered instabilities, which rely on the behaviour of CRs in the 'collapsed' streaming limit, such as the CR 'staircase' Quataert et al. (2022a), Huang & Davis (2022), and Hin Navin Tsung et al. (2022) in these simulations, as they have thus far been studied only in idealized setups. We intentionally include all the coupling terms necessary and the resolution requirement noted in e.g. Huang & Davis (2022) of $\Delta x \lesssim \kappa_\parallel/c_s \sim 8\,{\rm kpc}\,(T_{\rm gas}/10^6\,{\rm K})^{-1/2}$ (for observationally favoured κ at ~ 1 GeV) is easily satisfied, but as noted therein the instability depends on the plasma-β (but our experiments in described in Section 4.4.1 and Paper I vary this by factors of $\sim 10^4$). For now, we note that this behaviour does not appear to change any of our conclusions, nor did we expect it to do so, as (1) it only appears in the Alfvén-streaming (collapsed) limit; (2) in that limit if manifest in the ISM/inner CGM, it would not change the fact that the CRs have over-long residence times with $\delta_s \leq 0$; and (3) as a result the primary regime of interest for such behaviours is in the outer CGM (where more interesting observable effects for CR-driven outflows could be present), not the ISM.

As a result, either 'default' pure-ET-only (Alfvénic/GS95 or magnetosonic/YL04) model predicts CR scattering rates that are so low that one sees negligible secondary production at any energy, far-too-low a normalization of the CR spectra, etc. – the results are similar to those from the 'free escape' or 'low-start' model shown in Fig. 3. This is also shown explicitly around $\sim$ 1 GV in Hopkins et al. (2021d).

So instead, to give ET models the best possible chance of reproducing observations, in Fig. 4, we do not show the 'most theoretically favoured' ET models with their default normalization of the ET scattering rate $S_{\rm et,\pm}$, but instead allow the normalization of the scattering and damping rates to be free parameters. These normalizations are then rescaled to attempt to find a 'best fit' to observations.

In Alfvénic ET models, as described in Section 2.7 and 3.2, for any type of MHD/Alfvénic turbulence that obeys a critical balance-type assumption, the ET driving term must have the form $S_{\rm et,\pm} = \alpha_t\, e_{\rm B}\, \Gamma_{\rm turb}/(k_\parallel\, \ell_A)$. In the 'theoretically favoured' model, where one attempts to calculate α_t self-consistently from the same GS95-type MHD turbulence model as used for the cascade itself, one predicts $\alpha_t \ll 1$ (as small as $\sim 10^{-6}$–10^{-3}; see Chandran 2000; Lazarian 2016). Instead treating α_t as a free parameter, we find that, in order to approach roughly the correct order-of-magnitude normalization of the CR spectra and B/C ratios at $\sim$1–10 GV, we require $\alpha_t \sim 1$. But even then, if we include the standard damping terms (e.g. ion-neutral, NLL, dust), the cascade can be 'over-damped,' and still produce poor agreement with observations. So, to give the best-possible chance of reproducing observations (and also to highlight the 'pure ET' prediction), we ignore any damping other than the cascade transfer itself $\Gamma_{\rm turb}$. In other words, we have essentially assumed a pure, undamped, Alfvénic cascade, with arbitrary fitted normalization, so the *only* constraint on this ET model is the functional dependence on $k_\parallel$ that is *required* by critical balance.

Alternatively, we can consider a YL04-like magnetosonic model, which assumes the inertial-range cascade is isotropic, which is possible for e.g. fast modes on scales larger than the turbulent dissipation scales. But, this must account for the fact that, at $\lesssim$ TeV energies, the dissipation/Kolmogorov scale for magnetosonic modes is orders of magnitude larger than the gyroresonant scale. The 'theoretically favoured' version of this model is again over-damped (giving much-too-low $\bar{\nu}_s$), because any appreciable regions of the ISM that have neutral fractions $\gtrsim 10^{-3}$ or plasma $\beta_{\rm plasma} > 1$ produce a superexponential suppression of the scattering term $S_{\rm et,\pm}$ in this model. So again, to give the best-possible chance to reproduce observations, we follow Hopkins et al. (2021d) and ignore any ion-neutral or dust-damping and calculate the scattering and damping rates everywhere assuming $\beta_{\rm plasma} < 1$ (regardless of the real value of **B**). We use the full integral expressions from YL04 in the simulations, but for reference, this gives an approximate scattering rate $\bar{\nu}_s \sim c/(3\,\ell_A\, f_{\rm turb})$ where $f_{\rm turb} \sim {\rm MIN}[0.04\, c_s/v_{A,\,\rm ideal},\, \mathcal{M}_A\,(\nu_v/v_{A,\,\rm ideal}\,\ell_A)^{1/3}\,(k_\parallel\,\ell_A)^{1/6}]$, ν_v is the kinematic viscosity, and $\mathcal{M}_A$ is the Alfvénic Mach number of the turbulence at the driving scale. This is independent of rigidity.

We note these two models are akin to the 'Alfvén-Max' and 'Fast-Max' models studied in 'single-bin' CR models in Hopkins et al. (2021d), where we extensively varied the normalization and damping terms to try and fit the observed grammage as accurately as possible for $\sim 1-10$ GV protons (see also Section 4.4.1). We recover similar conclusions here for those rigidities.

However, we see in Fig. 4 that even if we freely renormalize the scattering and/or damping rates in these models to fit the proton spectra and secondary-to-primary ratios as best as possible at $\sim 1-10$ GV, there is a much bigger problem: both models qualitatively fail to

produce the observed dependence of B/C on rigidity, or the correct CR spectral shapes. This is because, as discussed in Section 3.2, at a fundamental level, if we allow for anisotropy/critical balance in Alfvénic models (even ignoring damping) or allow for finite damping/dissipation scales in magnetosonic models (even ignoring possible anisotropy and some of the more severe damping terms), this implies $\delta_s \leq 0$. In other words, the scattering rate cannot decrease as a function of CR rigidity as required by observations at $\gtrsim 0.1-1$ GV.[16]

For completeness, we have also run simulations assuming an undamped, isotropic, $\mathcal{E}(k) \propto k^{-3/2}$ cascade (chosen to have roughly the correct δ_s) across *all* energies and wavenumbers (i.e. ignoring all anisotropy terms, and all damping terms, and all SC terms, at all scales). This trivially gives $e_A \sim e_B\,(k_\parallel\,\ell_A)^{-1/2}$. But as noted in Section 3.2, this is not only unphysical but gives CR scattering rates a factor ∼1000 too large at all energies, vastly overpredicting e.g. secondary-to-primary ratios. We discuss models of this variety further below.

Given how widely we vary the amplitudes and damping rates and spectral indices of the ET models above, it should ultimately come as no surprise that subtleties such as the difference in the simulation-resolved properties of turbulence around the driving scale (e.g. the locally varying values of e_B and ℓ_A, or equivalently local $\mathcal{M}_A$) between different MW-like-simulated galaxies, different resolution levels, different initial **B**-field strengths, and other variations in Section 4.4.1 make no difference to our conclusions. Even if we ignore any of the resolved turbulence structure and simply assume a spatially universal $\mathcal{M}_A$, we obtain the same results (which again, is expected, given that our simple analytic toy model from Section 3.2 predicts the same discrepancies with observations).

5.2 Alternative damping requires discarding other damping models

We now consider the 'alternative damping' model from Section 3.5.1 and 4.4.2, with two examples illustrated in Fig. 5. First, we simply replace the 'standard' linear damping mechanisms ($\Gamma_{\rm in}+\Gamma_{\rm dust}+\Gamma_{\rm turb/LL}+\Gamma_{\rm nll,\pm}$) with a new $\Gamma_{\rm new,\,damp,\pm} \propto e'_{\rm cr}$. We use a best-fitting normalization of the variants we have explored, which is $\Gamma_{\rm new,\,damp,\pm} \sim (v_{A,\,\rm ideal}/\ell_A)\,(k_\parallel\,\ell_A)^{\xi_k}\,(e'_{\rm cr}/e_B)^{\xi_{\rm cr}}$ with $0.1 \lesssim \xi_k \lesssim 0.4$ and $\xi_{\rm cr}=1$. This has the desired effect, discussed in Section 3.5.1, of cancelling the $e'_{\rm cr}$ dependence in the SC driving term $S_{\rm sc,\pm}$, which is responsible for the instability/solution collapse problems (see Appendix A). Thus we can obtain a stable result in at least qualitative agreement with the observed behaviour at all rigidities, and *independent* of the CR energy density in the ICs (i.e. we converge to the same answer for 'low' and 'high' start ICs).

However, the challenge with this model is ensuring that $\Gamma_{\rm new,\,damp,\pm}$ dominates over other terms [specifically other damping terms) in the $D_t e_\pm$ equation (equation (7)], the balance of which set the equilibrium value of $e_\pm$ in the volume-filling ISM. Among the other standard terms in equation (7), we can retain or remove the 'gradient terms' (i.e. the 'advective' term $\nabla\cdot(v_{A,\pm}\,e_\pm\,\hat{\mathbf{b}})$ and 'PdV' term $(e_\pm/2)\,\nabla\cdot\mathbf{u}_{\rm gas}$), and/or the NLL damping term ($\Gamma_{\rm nll,\pm}$), and/or the 'default' (theoretically favoured, but weak) ET driving term $S_{\rm et,\pm}$, without qualitatively changing the behaviour seen in the top panels of Fig. 5. But unless we artificially remove or suppress the standard turbulent/linear Landau ($\Gamma_{\rm turb/LL}$), dust ($\Gamma_{\rm dust}$), and ion-neutral damping ($\Gamma_{\rm in}$) terms, they tend to dominate $\Gamma_\pm$ (e.g. $\Gamma_{\rm turb/LL} \gg \Gamma_{\rm new,\,damp}$). This causes the total damping $\Gamma_\pm$ to once again be dominated by terms that are independent of $e'_{\rm cr}$, and the 'solution collapse' problem returns. Similarly, we cannot simply increase $\Gamma_{\rm new,\,damp}$ until it dominates over all the other damping mechanisms at all CR energies: even though this will cure the instability, it will necessarily over-damp the scattering modes, producing too-low CR spectra and secondary abundances. One example of this failure is shown in Fig. 5. This illustrates that the discrepancy is not small – it would require more than order-of-magnitude changes in the expected strengths of turbulent, dust, and ion-neutral damping for typical MW conditions.

In summary, while a version of this model that can reproduce observations does exist, it requires a radical revision to our understanding of damping mechanisms. Not only one must introduce a novel damping mechanism with the desired $e'_{\rm cr}$ dependence, but also argue that the standard turbulent/linear Landau, dust, and ion-neutral (in diffuse but partially ionized phases) damping mechanisms are much weaker than currently understood, in order for this new damping mechanism to dominate with the correct normalization at all relevant rigidities.

5.3 Alternative sources

Figs 6 and 7 now consider the 'alternative driving' or 'alternative sources' models discussed in Sections 3.5.2 and 4.4.2.

5.3.1 Local/Linear source terms

First, in Fig. 6, we consider adding an alternative linear driving/source ($\zeta_A=1$ or $S_{\rm new}\propto e_A$) term. We take the form $S_{\rm new,\pm}=S_{\rm new,\,lin}\equiv\Psi_{\rm new}\,e_A$ with $\Psi_{\rm new}\equiv\Psi_0\,(k_\parallel/k_0)^{\zeta_k}$ (where we set $k_0\equiv$ au^{-1} for convenience without loss of generality). Because this is a linear driving term, the 'net' linear driving + damping $S_{\rm new,\,lin}-Q_\pm=(\Psi_{\rm new}-\Gamma_\pm)\,e_A$ is only weakly influenced by $S_{\rm new,\,lin}$ if $\Psi_{\rm new}\lesssim\Gamma_\pm$. So, for an initial experiment, we ignore the turbulent/linear Landau and dust damping mechanisms. In Fig. 6, we show that this form can give a plausible fit to the observed spectra and ratios for $0.6\lesssim\zeta_k\lesssim 0.9$. The required normalization is modest, e.g. $\Psi_0\sim 10^{-12}\,{\rm s^{-1}}\,(1+\mathcal{M}_A)\,(v_{\rm fast}/10\,{\rm km\,s^{-1}})\sim\delta v_{\rm turb}/{\rm pc}$ (or even $\Psi_0\sim 10^{-12}\,{\rm s^{-1}}\sim$constant), or similarly $S_{\rm new,\,lin}\sim 10^{-5}\,k_\parallel\,v_{\rm fast}\,(k_\parallel\,{\rm au})^{-1/3}$. In other words, the driving/growth rate favoured can be as little as ∼10^{-5} of the fast mode crossing rate.

The obvious challenge here, akin to the alternative damping discussed above (Section 5.2), is ensuring $\Psi_{\rm new}\gtrsim\Gamma_\pm$. Going through all terms in equation (7), the effects of $S_{\rm new,\,lin}$ are robust to retaining or removing the 'gradient terms,' or the other default source terms ($S_{\rm sc,\pm}$, $S_{\rm et,\pm}$), or the ion-neutral damping term ($\Gamma_{\rm in}$),[17] as well as

[16] At sufficiently low CR energies $\lesssim$ 100 MeV, it is notable in Figs 2–4 that even some models which produce qualitatively incorrect δ_s and qualitatively incorrect behaviours at higher energies can reproduce the spectral shapes and secondary-to-primary ratios of some species. This is because, as shown explicitly in Paper I, at these very low energies the rapidly increasing rate of Coulomb and ionization losses means that the residence time (at least in the disc mid-plane) can actually determined by the CR loss time-scales, and thus becomes independent from the predicted scattering rates $\bar{\nu}_s$.

[17] Unlike in the 'modified damping' case ($\Gamma_{\rm new,\,damp,\pm}$; Section 5.2), it appears that while $\Gamma_{\rm in}\gtrsim\Psi_{\rm new}$ in dense neutral ISM phases (CNM, molecular), which have a low volume-filling fraction and therefore do not strongly alter CR spectra in diffuse gas, we generally have $\Gamma_{\rm in}\lesssim\Psi_{\rm new}$ in warmer and/or more diffuse phases, even if they are partially neutral. This is especially true if we adopt a version of $\Psi_{\rm new}$ that scales with $\mathcal{M}_A$ or $\delta v_{\rm turb}$, which is larger in warm or cool phases.

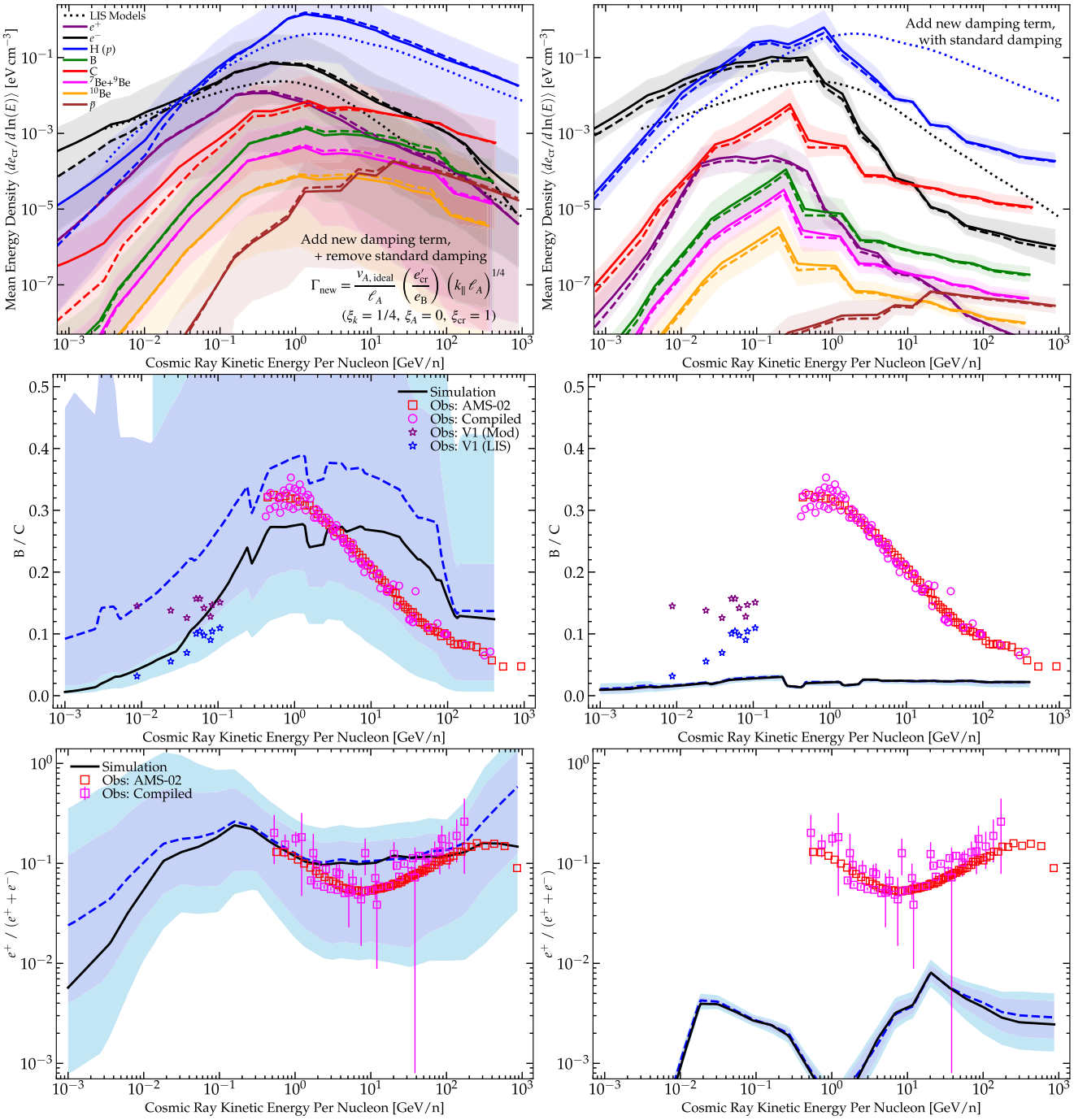

Figure 5. As Fig. 3, for models where we attempt to modify the damping physics to reproduce observations (Section 5.2). *Left:* a model where we take the default SC model, remove other ET sources ($S_{\rm et,\,\pm} \rightarrow 0$), and replace all the standard known linear damping terms (Section 2.5) with a damping term that scales as $\Gamma_{\rm new,\,damp} \sim (v_{A,\,\rm ideal}/\ell_A)\,(k_\parallel \ell_A)^{1/4}\,(e'_{\rm cr}/e_{\rm B})$, per Section 3.5.1. This depends on $e'_{\rm cr}$ in a way that cancels the term in SC driving which gives rise to the 'solution collapse' problems, and allows for a reasonable and *stable* solution (independent of the CR energy density in the ICs). *Right:* results if we retain this new damping term $\Gamma_{\rm new,\,damp}$, but re-introduce the ET driving and standard linear damping terms from e.g. ion-neutral, linear Landau, turbulence, and dust. Any of those linear damping terms is significantly larger than $\Gamma_{\rm new,\,damp}$ (for the normalization of $\Gamma_{\rm new,\,damp}$ needed to get a reasonable scattering rate) and 'swamps' it, producing results closer to our 'default' SC behaviour in Fig. 2, unless we make $\Gamma_{\rm new,\,damp}$ so large that the SC models are over-damped (giving $\bar{v}_{\rm s} \rightarrow 0$, so CRs are unconfined).

retaining or modifying/expanding the non-linear damping terms ($\Gamma_{\rm nll,\,\pm}$). However, if we do include the standard turbulent/linear Landau ($\Gamma_{\rm turb/LL}$) or dust ($\Gamma_{\rm dust}$) damping terms in their 'default' forms, these cause $\Gamma_\pm \gtrsim \Psi_{\rm new}$ in the volume-filling ISM, negating the effects of our added source term $S_{\rm new,\,lin}$.

Thus, while not totally implausible, this model does have theoretical challenges in dealing with the turbulent/linear-Landau and dust damping terms, akin to the modified damping scenario (Section 5.2). As discussed in Section 3.5.2, the advantage of this scenario is that it is quite easy to imagine linear instabilities operating on these

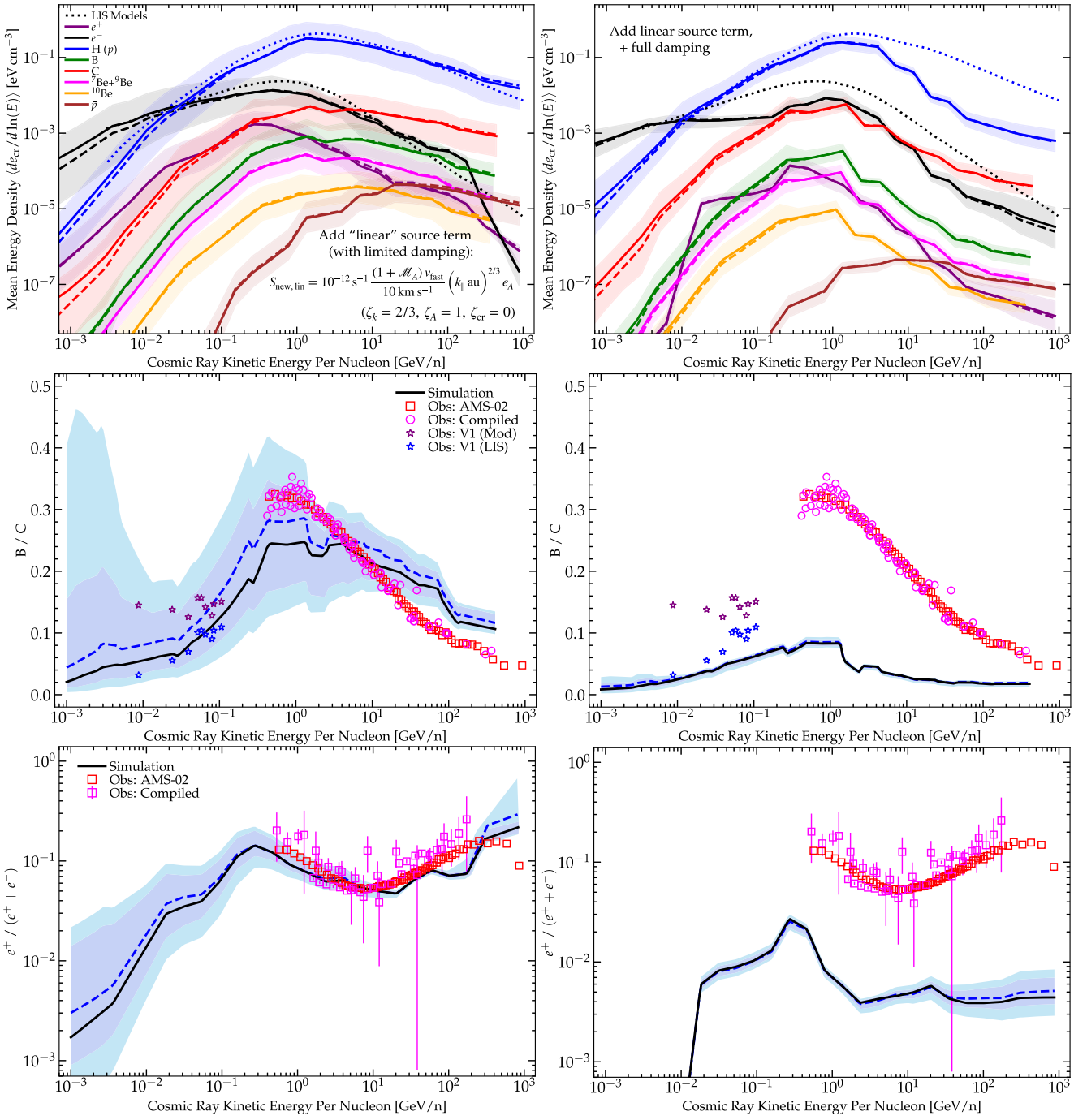

Figure 6. As Fig. 3, for models where we consider alternative linear scattering-mode driving/source terms ($S_{\rm new,\,lin} \propto e_A$, i.e. $\zeta_A = 1$; Section 5.3.1). *Left:* a model where we add linear source term $S_{\rm new,\,lin} \sim \Psi_{\rm new}\,e_A$ with $\Psi_{\rm new} \sim 10^{-12}\,{\rm s}^{-1}\,(k_\parallel\,{\rm au})^{2/3}\,(1+\mathcal{M}_A)\,(v_{\rm fast}/10\,{\rm km\,s^{-1}})$, removing specifically the turbulent/linear Landau and dust damping terms (keeping all other damping), and reducing the standard SC driving term by a factor ∼0.01. Note the agreement with observed ^{10}Be/^{9}Be and $\bar{p}/p$ (not shown) is also good. We also obtain broadly similar results for a simpler model with $\Psi_{\rm new} 10^{-12}\,{\rm s}^{-1}\,(k_\parallel\,{\rm au})^{2/3}$, but the fit is not quite as good (this leads to flatter B/C at high-energies). *Right:* as *left* but re-introducing the dust and turbulent/linear Landau damping terms, which are usually larger than the linear growth term, so the behaviour reverts to be closer to the 'default' SC model and collapses at intermediate and high-energies to the unconfined solutions.

scales with roughly the correct growth rate and $k_\parallel$-dependence. For example, multifluid instabilities like the Kelvin–Helmholtz instability would have growth rates $\sim (\delta\rho/\rho)\,k\,\delta v$, so would only require $(\delta\rho/\rho)\,(\delta v/v_{\rm fast}) \sim 10^{-5}$ on these scales to grow at roughly the correct rate. RDIs in the 'mid-k' range, which may be applicable at these scales, and RTIs similarly have growth rates of $\sim \sqrt{a\,k}$ where a is some differential acceleration between e.g. dust and gas (for RDIs) or a fluid interface (for the RTI). Given the low normalizations, even a very small differential acceleration $a \sim 10^{-4}\,v_A^2/\ell_A$ could be sufficient to drive the required growth rates. Of course, for 'interface' instabilities one must ask what the interface would be, while for RDIs, the 'high-k' modes often have a less-favoured scaling $\propto k^{1/3}$, and more

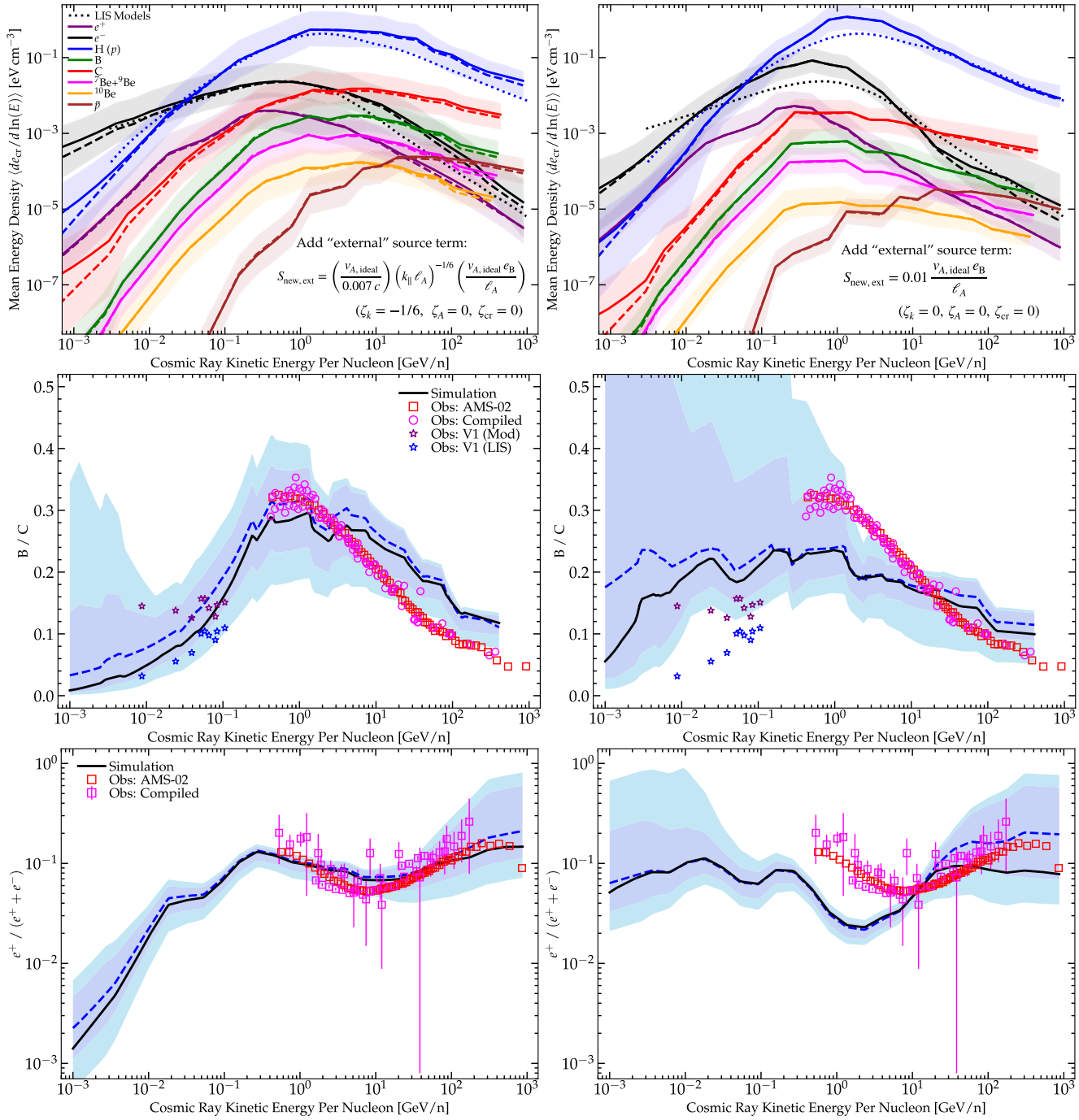

Figure 7. As Fig. 3, for models where we consider alternative fixed or 'external' (e_A-independent, $\zeta_A = 0$) scattering-mode driving terms ($S_{new,ext}$; Section 5.3.2). Here, we *do not* disable any of the standard damping or other driving terms, we simply add this additional source term. *Left:* a model with $S_{new,ext} = (v_{A,ideal}/0.007\,c)\,(v_{A,ideal}/\ell_A)\,(k_\parallel \ell_A)^{-1/6}\,e_B$ (we also obtain similar results for $S_{new,ext} \sim 0.005\,(v_{A,ideal}/\ell_A)\,e_B\,(k_\parallel \ell_A)^{-1/6}$). Adding a weakly scale-dependent driving term of this form ($\zeta_k \approx -1/6$, $\zeta_A = \zeta_{cr} = 0$), with amplitude comparable to ~ 1 per cent of turbulent or Alfvén dissipation rates can produce reasonable behaviours, without having to strongly modify known damping or other driving terms. Note agreement with ^{10}Be/^{9}Be and $\bar{p}/p$ is good as well. *Right:* a model with $S_{new,ext} = 0.01\,(v_{A,ideal}/\ell_A)\,e_B$ (roughly $\sim 0.01\,\rho\,\delta v_{turb}^3/\ell_{turb}$ on the driving/resolved ISM scales). This has similar amplitude and behaviour but slightly different wavelength-dependence ($\zeta_k = 0$). While it does not fail catastrophically, the agreement with observations is notably worse, demonstrating that the favoured range of $-0.25 \lesssim \zeta_k \lesssim -0.1$ is relatively constrained.

extreme conditions and/or certain modes (e.g. the 'CR like' RDI modes) could produce over-confinement (see Squire et al. 2021 and Ji et al., in preparation). It is even conceivable that $S_{new,lin}$ could arise at CR energies $\gg$ GV from the action of the Bell (2004) instability sourced by the dominant $\sim$ GV CRs (i.e. the long-wavelength regime of the instability sourced by lower-energy CRs); however, this could introduce some (but not all) of the 'instability' problems from SC (Section 3.4).

In these regimes, it is also not implausible to assume that the usual $\Gamma_{turb/LL}$ and Γ_{dust} terms would be strongly modified. The expressions

for 'turbulent/linear Landau' damping reviewed in Section 2.5 are derived specifically assuming that the modes are sheared out by *external* turbulence from a standard GS95 cascade from larger scales, where the dominant driving of any modes that are *not* exactly parallel modes driven by SC comes from that ET cascade (see Yan & Lazarian 2002; Farmer & Goldreich 2004; Zweibel 2017). But if $S_{\rm new,\,lin}$ dominates over $S_{\rm sc,\,\pm}$ and $S_{\rm et,\,\pm}$, then some of these assumptions will not apply – so there is not necessarily any reason to think the modes would be sheared out in this manner. Similarly, the dust damping rate $\Gamma_{\rm dust}$ is derived specifically under the assumption that the RDIs are negligibly weak/inactive (indeed the damping and instability arise from similar physical effects) – if there is sufficient dust drift to cause an RDI, the dust switches from being a 'damping' to a 'driving' term (see Squire et al. 2021).

5.3.2 *External/fixed-rate source terms*

We now consider adding a 'constant' or 'external' alternative driving/source term ($\zeta_A = 0$ or $S_{\rm new} \propto e_A^0$), in Fig. 7. This is the most straightforward successful model variant we consider. Keeping everything else in our 'reference' model fixed, we can simply add a source term $S_{\rm new,\,ext} \propto k_\parallel^{\zeta_k}$ (where we find best fits with $-0.25 \lesssim \zeta_k \lesssim -0.1$), and normalization e.g. $S_{\rm new,\,ext} \sim 0.005\,(v_{A,\,\rm ideal}/\ell_A)\,(k_\parallel\,\ell_A)^{-1/6}\,e_{\rm B}$ or $\sim 0.01\,\delta v_{\rm turb}^3/\ell_A$, i.e. ~ 1 per cent of the typical turbulent dissipation rates. This produces remarkably good behaviour across all diagnostics we consider.

While a detailed exploration is outside the scope of this work, we have also applied the analysis pipeline from Paper I to this model to explore its predictions for the spatially resoled γ-ray emissivities and spectra and CR ionization rates in the Galaxy (since any successful model must reproduce these, as well) and we find agreement there as well, broadly similar to the favoured phenomenological model from Paper I which is shown in Fig. 1.

Unlike the alternate linear-damping or driving models, we do not have to 'remove' or re-tune any of the known terms (SC or ET driving or different damping mechanisms) to see good behaviour here. In other words, this model works with no other unwarranted modifications to the wave-damping or source physics. The one caveat is the SC driving in 'high-start' IC cases. Adding $S_{\rm new,\,ext}$ with $\zeta_A = 0$ prevents the SC instability from collapsing to the 'free-streaming' solution, because $S_{\rm new,\,ext}$ sets a minimum driving even if $e'_{\rm cr}$, and hence $S_{\rm sc,\,\pm}$, is low. But if $e'_{\rm cr}$ is sufficiently high and we still include our standard $S_{\rm sc,\,\pm}$ in $S_\pm$, we can have $S_{\rm sc,\,\pm} \gtrsim S_{\rm new,\,ext}$, with $S_{\rm sc,\,\pm}$ large enough to push the system into the 'infinite-confinement' branch of solution collapse (so that this causes $e'_{\rm cr}$ and $S_{\rm sc,\,\pm}$ to continue to rise). This scenario cannot be halted by the added $S_{\rm new,\,ext}$ term, and indeed does still occur if we just add $S_{\rm new,\,ext}$ in 'high start' IC simulations. While it is plausible that such collapse could occur physically in extreme regions – e.g. galactic nuclei, or starburst galaxies, which are observed to be at the proton calorimetric limit in γ-ray emission; see Lacki et al. 2011; Tang, Wang & Tam 2014; Griffin, Dai & Thompson 2016; Wojaczyński & Niedźwiecki 2017; Wang & Fields 2018) – it obviously does not occur for typical MW conditions. So to ensure it does not occur, we find that our results are most stable if we reduce $S_{\rm sc,\,\pm}$ by a factor $\sim$10–100 from its 'reference' value. But as discussed above in Section 3.4 and in e.g. section 5.3.4 of Hopkins et al. (2021d), such a renormalization of $S_{\rm sc,\,\pm}$ is plausible, based on corrections to $S_{\rm sc}$ from more careful detailed PIC modelling of pitch-angle dependence, helicity, non-linear, and non-gyroresonant effects (e.g. Bai et al. 2019; Holcomb & Spitkovsky 2019). Of course, these would need to be revisited in more realistic situations with some $S_{\rm new,\,ext}$ term present.

It is also noteworthy that this functional dependence of $S_{\rm new,\,ext}$ on k, e_A, and $e_{\rm cr}$ (e.g. $\zeta_A = 0$, $\zeta_{\rm cr} = 0$, and $|\zeta_k|$ small) is superficially similar to what one would obtain in the simplest 'classical' isotropic, undamped, inertial-range K41-like turbulent cascade, where $S_{\rm et,\,\pm} \sim$ constant is the turbulent dissipation rate. This, plus the fact that the dimensional dependence of $\Gamma_{\rm turb/LL}$ and $\Gamma_{\rm dust}$ on $k_\parallel$ are similar to the turbulent cascade rate, is indeed why, as many have noted previously, the observed $\delta_{\rm s}$ is not so different from what one would naively obtain from a 'traditional' isotropic undamped ET model with a spectrum similar to $\mathcal{E}(k) \propto k^{-3/2}$ ((neglecting dissipation, anisotropy, and finite dynamic-range effects; see e.g. discussion in Blasi & Amato 2012; Vladimirov et al. 2012; Gaggero et al. 2015; Cummings et al. 2016; Guo et al. 2016; Jóhannesson et al. 2016; Korsmeier & Cuoco 2016; Evoli et al. 2017; Amato & Blasi 2018). But there are fundamental physical differences here. Most importantly, as argued above and in Appendix C in detail, this $S_{\rm new,\,ext}$ cannot stem from a traditional undamped Alfvénic or magnetosonic cascade from large ISM scales. All of the effects reviewed therein would prevent $S_{\rm new,\,ext}$ from having the form assumed. It is possible that some sort of 'mini-cascade' could occur over a small range of scales, with smaller-scale driving, provided it could avoid the anisotropy and damping problems we have outlined. But as justified formally in Appendix C, we easily avoid all of these conceptual difficulties if we simply assume $S_{\rm new,\,ext}$ represents driving of Alfvénic modes competing directly with damping *at each scale* separately – we are simply arguing for a driving mechanism whose power is only weakly scale-dependent. Such an effect could possibly arise, for example, if reconnection played an important role in MHD turbulence at small scales. Such a scenario would require that flux ropes formed by reconnection between sheets in the perpendicular plane (Schekochihin 2022) subsequently broke up in the parallel direction with the right spectrum, which is plausible but highly speculative. However, it is worth emphasizing that since the required power in $S_{\rm new,\,ext}$ is two or three orders of magnitude smaller than the power in the turbulent cascade, these fluctuations should be strongly subdominant and would be very difficult to observe in simulations. Finally, we note that the true best-fitting driving favours a modest scale-dependence $-0.25 \lesssim \zeta_k \lesssim -0.1$ (cf. left-hand and right-hand panels of Fig. 7); this is not steep, but is distinctly different from the predictions of any turbulence models in the literature.

5.3.3 *Summary of requirements*

We can summarize the required scaling for a viable driving/source term $S_{\rm new}$ for 'linear' $S_{\rm new,\,lin}$ (Section 5.3.1) and 'external' $S_{\rm new,\,ext}$ (Section 5.3.2) cases as follows:

$$S_{\rm new,\,lin} \sim 10^{-12}\,{\rm s^{-1}}\,e_A\left(\frac{k_\parallel}{\rm au^{-1}}\right)^{\zeta_k} f_{\rm S}(...)\,\left(0.6 \lesssim \zeta_k \lesssim 0.9\right)$$
$$S_{\rm new,\,ext} \sim 0.01\,\frac{v_{A,\,\rm ideal}}{\ell_A}\,e_{\rm B}\left(\frac{k_\parallel}{\rm au^{-1}}\right)^{\zeta_k} f_{\rm S}(...)\,\left(-0.25 \lesssim \zeta_k \lesssim -0.1\right) \quad (13)$$

where $f_{\rm S}(...)$ is some function of ISM/plasma properties. Any viable driving mechanism must therefore satisfy the following conditions: **(1)** It must drive modes of interest, i.e. Alfvénic modes with $k_\parallel$ in

the relevant range and that are not too-extreme in their anisotropy.[18] We stress that equation (13) refers to the driving rate of these modes, specifically, not to other (e.g. nearly perpendicular) modes, which are generically less efficient scatterers and would require a larger $S_{\rm new}$. **(2)** It must be able to drive modes across the wavelength scales of interest. For rigidities $R \sim 0.001 - 1000\,{\rm GV}$ studied here, this is $1/k_{\parallel} \sim r_{g,{\rm cr}} \sim 3\,B_{\rm \mu G}^{-1} \times 10^{9-15}\,{\rm cm}$. However, it is possible that very low-energy CRs ($\lesssim 100\,{\rm MeV}$) have residence times that are primarily regulated by ionization/Coulomb losses (as argued empirically in Hopkins et al. 2021a and found in some of our experiments), which would increase the lower limit to $\sim 10^{11.5}\,B_{\rm \mu G}^{-1}\,{\rm cm}$. Similarly, it is plausible that gyroradii approach/exceed the dissipation scales of fast magnetosonic turbulence (so 'traditional' ET theory becomes viable) above the scales relevant to few-hundred GV CRs (e.g. Fornieri et al. 2021), in which case the upper limit could decrease to $\sim (0.3-1)\times 10^{15}\,B_{\rm \mu G}^{-1}\,{\rm cm}$. **(3)** It must have one of the forms above in equation (13), with the range of ζ_k corresponding to the extrinsic or linear driving ($\zeta_A = 1$ or $=0$), with $f_{\rm S}(...)$ parametrizing all the dependence on the ISM plasma physics. **(4)** In order to match the normalization in equation (13), the appropriate volume or scattering-rate weighted average $\langle f_{\rm S}(...)\rangle$ (parametrized in the same way) must be $\sim$1 integrated from CR sources to the Solar LISM in a MW-like galaxy through most of the volume-filling ISM. **(5)** By definition, $f_{\rm S}$ must depend weakly or not at all on CR properties (e.g. the DF f, number density $n_{\rm cr}$, energy density $e_{\rm cr}$, streaming speed $v_{\rm st}$, etc.); weakly or not at all on $k_{\parallel}$ (such that $S_{\rm new}$ has the correct $k_{\parallel}$ dependence parametrized by the range of ζ_k); and weakly or not at all on the local mode amplitude e_A or $\delta\mathbf{B}(k_{\parallel})$ (i.e. the driver has $\zeta_A \approx 0$ or ≈ 1, appropriately).

Briefly, it is worth noting that the favoured ranges of ζ_k for these driving mechanisms (or the alternative damping in Section 5.2) in our simulations are slightly different from that analytically anticipated from our simple steady-state back-of-the-envelope calculations in Sections 3.5.1–3.5.2. This is not surprising: in our simple model, we neglected losses, adiabatic terms, contributions to transport from Alfvénic streaming, the interplay of multiple damping/source mechanisms, and finite source/scattering halo distributions, all of which contribute some additional rigidity-dependence to the final behaviour, in a way that only our full simulations can accurately capture. But crucially, the qualitative behaviours and conclusions are identical, with only modest quantitative corrections. This suggests that the general physical principles are robust.

5.4 Can different galaxy properties rescue SC or ET models?

It is natural to ask whether there might be some different galaxy properties (perhaps some difference between the real MW and our models or assumptions here) that could resolve the discrepancies with observations, without invoking new driving or damping mechanisms. We have attempted to explore this with both our general analytic arguments and, to the extent possible in our simulations, with the parameter variations discussed in Section 4.4.1. Specifically, for the general SC and ET models in Figs 2–4, we have run simulations using three different cosmologically selected MW mass galaxies, which – while all selected to have properties that are broadly similar to the MW – differ in detail (e.g. different sizes, gas density and star formation rate distributions, presence or absence of bars and arms, etc.). We have also arbitrarily renormalized the initial magnetic fields and CR energy densities in the simulations by large factors as discussed above. And for all of our simulations, we have a large number of independent snapshots sampling several galaxy dynamical times – we have checked to confirm that the results in our figures are robust (approximately steady-state) in time, and to see whether there could be even a transient phase where the SC and ET models produce good simultaneous agreement with different observables. As relates to all these differences (variations in time, between different MW mass galaxies, or between modified ICs), our key conclusions are robust. Indeed, the differences between galaxies or different times are much smaller than the differences between models (see Paper I for more detailed comparisons).

However, it is not possible in computationally expensive simulations like ours to survey all possible galaxy properties. So one might ask whether there still exists some hypothetical combination of plasma parameters that would allow the SC and/or ET models to reproduce observations. This is essentially the question explored in Kempski & Quataert (2022), of which we became aware during the writing of this manuscript. While our experiments in this paper might be described as 'constraining which CR scattering models can reproduce observations, given a set of galaxy models,' Kempski & Quataert (2022) effectively consider the complementary question 'given a fixed CR scattering-rate model, what galaxy model could reproduce observations?'. Specifically, they consider analytically parametrized models of a stratified disc + CGM and show, in agreement with our conclusions, that neither SC nor ET models can possibly reproduce the observations alone.[19] However, they do argue that the combination SC+ET allows a match to observations, in principle, if the stratified disc + CGM follows a specific particular model. However, as Kempski & Quataert (2022) caution, this requires a very specific and fine-tuned set of assumptions: their model requires that the profile of the Alfvén speed, turbulence strength, ionization fraction, and $e'_{\rm cr}$ follows a specific profile as a function of scale-height. This allows ET driving with a scaling close to our modified 'Fast-Max' model in Fig. 4 to dominate within the thick disc (with a certain strength), while SC driving with non-linear-Landau damping and the 'collapsed' Alfvénic streaming solution only dominates outside the disc in the CGM (with that following a specific vertical Alfvén-speed profile). Essentially, in their model, the profiles of relevant plasma properties (like v_A), which appear in the scalings of $\bar{\nu}_{\rm s}$ for the SC and ET models, are chosen such that they 'cancel out' the fundamental problematic scalings of SC or ET alone.

We have attempted to explore a model akin to this best fit of Kempski & Quataert (2022), by (1) replacing all our driving and damping terms with just the combination of SC driving plus the 'Fast-Max' ET driving model (the same as the scalings adopted in Kempski & Quataert 2022), with just NLL damping, while also (2) renormalizing $\mathbf{B}$ and $e'_{\rm cr}$ in our ICs to match the vertical profile of $e'_{\rm cr}$ and v_A assumed therein. But we find this experiment quickly undergoes the same 'solution collapse' akin to our 'normal' or 'high' start ICs in Figs 2–3. The difference may be that it is simply not possible to exactly reproduce all of the assumptions of the analytic model in our ICs; e.g. because the galaxy density profile cannot be freely renormalized in our simulations, and/or because we include SC + ET terms together, while Kempski & Quataert

[18] As shown in Appendix C, the modes do not have to be specifically parallel or isotropic, but should at least obey $|k_{\parallel}| \gg (|\delta\mathbf{B}(k_{\parallel})|/|\mathbf{B}|)\,|k_{\perp}| \sim 0.0003\,R_{\rm GV}^{0.2}\,|k_{\perp}|$.

[19] See also Fornieri et al. (2021), who similarly concluded that ET models alone (Alfvénic or akin to our 'Fast-Max' $S_{\rm et,\pm}$) could not reproduce observations below a few hundred GV, even allowing for arbitrary freely fit galaxy/ISM properties in a parametrized analytic model.

(2022) consider a model where only one or the other contributes meaningfully at any given scale-height. But the bigger challenge may be that our simulations dynamically evolve quantities like **B** and $e'_{\rm cr}$, and these will quickly deviate from their initial values as e.g. super-bubbles and clustered SNe explode. These will then push the system away from equilibrium and into one of the solution-collapse regimes. This suggests, at least, that this fine-tuning is not trivial to achieve in practice, and is unlikely to be the case in the MW (as, indeed, is concluded by Kempski & Quataert 2022 also).

5.5 What about the 'meso-scale'?

In thinking about our conclusions, it is helpful to separate the enormous hierarchy of scales into three groups: 'macro,' 'micro,' and 'meso' scales. For our purposes, we can think of the 'macro-scale' structures as those which are at least semiresolved by our simulations (larger than a few thousand Solar masses). This includes e.g. the multiphase structures of the ISM, and clumping of gas (e.g. the existence of GMCs); global galactic structure (the nucleus, disc, and bulge, bars, and spiral arms); the scale heights of the cold, star-forming disc (and young stellar disc) and the warm/thick gas and stellar discs, and the associated driving scales of ISM turbulence; clumping of star formation and SNe (in space and time), and associated super-bubbles and galactic chimneys; galactic fountains and the ISM-CGM interface; the existence of a turbulent CGM and galactic outflows, and the interaction with satellite galaxy ISM/CGM structure. All of these, it is worth noting, have been extensively studied and compared to observations with simulations identical to those here (modulo the assumed CR scattering rate scalings; see references in Sections 1 and 4 and Chan et al. 2019; Hafen et al. 2019; Emami et al. 2019; Ji et al. 2020; Benincasa et al. 2020; Gurvich et al. 2020; Chan et al. 2021; Ponnada et al. 2022; Kim et al. 2022; Trapp et al. 2022). The point of our numerical simulations, fundamentally, was to see if non-linear effects from structure (e.g. varying values of terms which go into estimating scattering rates, such as $|\mathbf{B}|$ or n, as we discussed immediately above in Section 5.4) could somehow introduce qualitatively different behaviours from those predicted by the simple analytic arguments in Section 3, and so somehow 'rescue' traditional SC/ET models from the problems we anticipated. We also wanted to explore whether macroscopic 'back-reaction' or CR 'feedback' effects, which should be resolveable, would somehow lead to a kind of feedback loop that could alter our analytic conclusions. This includes e.g. the effects of CRs changing galactic wind/outflow dynamics, driving instabilities such as the 'staircase,' or altering the phase structure of the CGM, or exerting 'pressure' to change the vertical balance or turbulent structure of the ISM, or altering the global ionization structure of cold clouds – all of these can (and as noted above, many do to some extent) occur in our simulations, which include all of the required physics and coupling terms.

Of course, the simulations have finite resolution and as we clearly noted from the beginning of this study, they cannot even approach resolving the 'micro-scale' by which we refer to gyroresonant scales for the CRs of interest ($\lesssim 100\,$au). These are the scales of actual 'scattering' physics, where PIC-type methods are needed to treat the CR dynamics. In the simulations, CR scattering is therefore explicitly 'subgrid.' Another way of saying this is that we cannot predict CR scattering rates from first principles, but instead are here testing different models for how the 'microscale' CR scattering rates depend on 'macroscale' parameters. This allows us to show that some key physics or assumptions must be missing from these models.

But it is also worth mentioning that, given this scale separation, there could be interesting dynamics in the 'meso-scale' as well, by which we mean scales much larger than the gyroresonant scales, but much smaller than the resolved simulation scales or driving/coherence scales of the volume-filling warm ISM/CGM components (and with small volume-filling factors). Consider, for example, stellar magnetospheres: we know from the Heliosphere that these represent regions with order-one changes in the magnetic field on scales $\sim 100\,$au, vastly smaller than the Alfvén/coherence scale of magnetic fields in the volume-filling warm ionized ISM ($\sim 100-200\,$pc), and that this can (and does) strongly scatter/deflect the pitch angles of CRs with energies $\ll$ TeV. In a sense, we can think of this as a tiny patch of the ISM interior to which the local Alfvén scale ℓ_A decreases from $\sim 100\,$pc to $\sim 100\,$au. These are obviously un-resolved in our simulations. But it is, at least in principle, possible to imagine models in which such 'meso-scale' structures dominate CR scattering, and introduce effects like those we sought to explore on the 'macro-scale' in our simulations, and so could strongly modify the 'effective' scattering rates and residence times of CRs. We stress that any such model would still represent a radical departure from traditional CR transport theory: in traditional models such as those explored here, scattering is dominated via the sum of many small-angle/weak scattering events, and the CR residence times (hence 'effective' scattering rates) are dominated by the statistically homogeneous, relatively smooth, volume-filling phases of the ISM (e.g. the WIM and warm inner CGM, for CRs observed in the LISM; see Paper I and references therein). And any such model would still have to solve the problems we present here: it would have to predict a physical means by which such structures could introduce any (let alone the *correct*) energy dependence to the 'effective' CR scattering rates over the required energy range. In this sense, one can think of such models as a mechanism by which something like our alternative damping or driving rates (required on macro-scales) could be achieved, just via an intermediate scale effect. But in addition, such a model would necessarily have to show that meso-scale structures actually dominate CR scattering between their initial acceleration and observation in the LISM. For the example of stellar magnetospheres given above, this appears impossible: the mean-free-path between magnetospheres in the ISM (given a stellar density of $\sim 1\,{\rm pc}^{-3}$ and radius $\sim 100\,$au) is $\sim$Mpc, while the observationally inferred mean-free-path for deflection/scattering of $\sim$GeV CRs is $\sim 10\,$pc (10^5 times smaller). And implicit in the above, it would still be necessary in such a model to explain how the diffuse ISM outside of such structures does not undergo solution collapse or SC runaway to over-confinement. Still, it is very much worth keeping such models in mind, as there is a diverse ensemble of phenomology in the ISM on scales not captured in either the simple analytic scalings or numerical galaxy-scale simulations explored here (for examples, see the scattering processes examined in Bai 2022; Beattie et al. 2022), and we will in future work (Butsky et al., in preparation) try to map out in more detail some of the requirements of such meso-scale models.

6 CONCLUSIONS

We have combined analytic models and a suite of detailed numerical simulations of CR transport in fully dynamical Galactic environments to explore the physics of CR scattering at CR energies ~MeV-TeV. From all of these, we show that standard SC and ET models cannot even qualitatively reproduce basic features of the observed CR spectra and secondary-to-primary or radioactive species ratios. The model failures are not superficial and, across our extensive survey, we

find no 'tweaking' that acts as a remedy; this is expected, because we argue that the problems arise due to fundamental and indeed *defining* assumptions of SC and ET models. Specifically, for SC models, the fact that the term driving the growth of the CR scattering rate itself depends on the CR flux or energy density causes the SC 'instability' or 'solution collapse' problem, wherein, regardless of any details of the functional form of SC scattering rates or damping mechanisms, CRs quickly converge to either the trapped/infinite-scattering limit or the free-streaming/escape-at-c limit. For ET models, the assumption that the scattering modes arise from an MHD 'cascade,' or other transfer between scales over a large dynamic range, forces the scattering modes to obey the qualitatively incorrect scaling as a function of rigidity at scales below the Alfvén and/or dissipation scale of turbulence (which includes all CRs in the ISM below a few hundred GeV).

We therefore phenomenologically approach the problem and ask 'what would be needed' – in terms of either the driving or damping of CR scattering modes – to resolve all of these issues and reproduce CR observations. While previous studies have empirically quantified this in terms of an 'effective diffusivity' or 'mean scattering rate' that best fits observations (e.g. fitting some constant in space and time or a simply parametrized function for the diffusion coefficient as a function of CR rigidity), we go further and actually solve the dynamical equations for the CR scattering rates, incorporating what is known about driving and damping rates of parallel magnetic fluctuations. For the first time, we constrain 'what is needed' directly in terms of the *local* driving rate $S_{\pm}$ or damping rate $\Gamma_{\pm}$ of CR scattering modes, on scales of order the gyroresonant wavelengths. These are the quantities that can actually be predicted by detailed theoretical calculations and PIC simulations of CR scattering physics. We identify three classes of model that could, at least qualitatively, reproduce the CR observations and quantify what is needed for each.

(i) Alternative Damping: All the key problems introduced by the dominant SC term at ∼MeV-TeV energies can be resolved if the linear damping rate scales with the CR energy density (at some rigidity) $\Gamma_{\rm new,\,damp,\,\pm} \propto e'_{\rm cr} \propto d e_{\rm cr}/d\ln R_{\rm cr}$, e.g. $\Gamma_{\rm new,\,damp,\,\pm} \sim (v_{A,\,\rm ideal}/\ell_A)\,(k_\parallel\,\ell_A)^{\xi_k}\,(e'_{\rm cr}/e_{\rm B})$ with $0.1 \lesssim \xi_k \lesssim 0.4$. However, there are two key issues: (1) it is not obvious what could physically produce such a scaling, and (2) this damping must dominate over all other linear damping mechanisms in the volume-filling ISM, which effectively requires discarding or drastically reducing the normalization of standard linear damping mechanisms such as ion-neutral, dust, turbulent/linear Landau, and NLL damping.

(ii) Alternative Linear Driving/Sources: Alternatively, if the CR scattering waves with energy $e_A \sim |\delta\mathbf{B}(k_\parallel)|^2$ are driven by a linear source term, $S_{\rm new,\,lin} \propto e_A$, where $S_{\rm new,\,lin}$ does *not* depend on CR energy, this can avoid the problems of SC models and reproduce observations. A form such as $S_{\rm new,\,lin} \sim 10^{-12}\,{\rm s^{-1}}\,e_A\,(k_\parallel\,{\rm au})^{\zeta_k}$ with $0.6 \lesssim \zeta_k \lesssim 0.9$ provides reasonable results. Note that SC models are intrinsically based on such a 'linear' source term (from CR gyroresonant instabilities), but the problem is that their dependence on the CR energy density introduces the instability/solution-collapse problems, and the k dependence scales incorrectly to reproduce observations. But a wide variety of other known linear instabilities – e.g. a host of multifluid instabilities that are known to operate on the relevant scales – could potentially explain this scaling, and only very modest power is needed in the relevant modes. The problem with this solution is that the linear source must compete with linear damping and SC driving, so reproducing observations with this model class requires somewhat weaker linear damping. This problem is not as severe as for 'alternative damping' above, but in particular the standard turbulent/linear-Landau and dust damping scalings are too strong and would need to be revised.

(iii) Alternative External Driving/Sources: Instead, an alternative source term that is independent of e_A and $e'_{\rm cr}$, and only weakly dependent on k – for example, $S_{\rm new,\,ext} \sim dE(k_\parallel)/d\ln k_\parallel\,dt\,d\,{\rm Volume} \sim 0.01\,(v_{A,\,\rm ideal}/\ell_A)\,e_{\rm B}\,(k_\parallel\,{\rm au})^{\zeta_k}$ with $-0.25 \lesssim \zeta_k \lesssim -0.1$ – can resolve the key problems of SC and ET models and reproduce observations. This version requires remarkably little revision to other known damping or driving terms. While this 'external' scaling is qualitatively similar to ET models in that $S_{\rm new,\,ext}$ is independent of $e'_{\rm cr}$ and e_A, it cannot derive from a standard turbulent cascade from large scales without introducing the anisotropy and damping problems, but is better thought of as a driver acting over a wide range of scales (or some modification of standard MHD turbulence paradigms). The total power needed is modest (∼ 1 per cent of the dissipation rate in ISM turbulence), and it is plausible to imagine a variety of physical mechanisms that could act in this way: the challenge may be to ensure such a mechanism can act over the entire relevant dynamic range of $\sim 10^3 - 10^6$ in $k_\parallel$.

It is important to note that, although we demonstrate the conclusions above over a wide range of CR energies ∼MeV-TeV, it may be possible to somewhat reduce the dynamic range of CR energies (and therefore wavenumbers $k_\parallel$) over which alternative physics must play a key role. For example, as argued in Paper I and seen in some (but not all) of the models here, the residence time of very low-energy CRs at $\lesssim 10 - 100$ MeV could be regulated by Coulomb/ionization losses (making predictions consistent with observations and nearly independent of scattering rates), so long as the scattering rates at these energies are sufficiently high so that the diffusion/escape time is longer than loss times. And depending on detailed ISM properties, at some energy $\gtrsim 0.1 - 1$ TeV, CR gyro radii will eventually become comparable to the dissipation scales of turbulence, so the 'classical' ET scenario of scattering from an undamped extrinsic turbulent cascade becomes a reasonable approximation (provided there is an isotropic fast-magnetosonic inertial-range cascade with approximately $k\,\mathcal{E}(k) \propto k^{-3/2}$ and the correct normalization). It is at intermediate energies, where most of the CR energy density resides, that the problems described here are most acute.

We stress that we are not here advocating for any one specific physical process as the explanation for CR scattering. Instead, our goal is to identify and further investigate generic problems with current SC and ET models; some of these problems were already known, some have been first identified here. We further identify classes of scalings for either driving or damping of CR scattering modes that could, in principle, explain the observations, discussing several possible physical mechanisms above. In future work, we hope to explore some of these candidate processes in more detail to assess if any can actually produce the correct scaling and normalizations needed to explain observations. It may be that the quantitative details will differ by a modest amount, as there are a variety of effects that lead to e.g. exact deviations from simple power-law behaviour, but we expect that the key qualitative requirements identified herein are robust. If mechanisms can be identified that meet these criteria, it will be important to also test them in microphysical MHD-PIC-like simulations, then use the results as input to galactic simulations such as those explored here. This will allow a quantitative comparison to CR observations, providing further valuable constraints on the important processes at play.

ACKNOWLEDGEMENTS

Support for PFH was provided by NSF Research Grants 1911233 & 20009234, NSF CAREER grant 1455342, NASA grants 80NSSC18K0562, *HST*-AR-15800.001-A. Support for JS was provided by Rutherford Discovery Fellowship RDF-U001804 and Marsden Fund grant UOO1727, which are managed through the Royal Society Te Apārangi. Numerical calculations were run on the Caltech compute cluster 'Wheeler,' allocations FTA-Hopkins/AST20016 supported by the NSF and TACC, and NASA HEC SMD-16-7592.

DATA AVAILABILITY STATEMENT

The data supporting this article are available on reasonable request to the corresponding author.

REFERENCES

Abdollahi S. et al., 2017, Phys. Rev. D, 95, 082007
Acero F. et al., 2016, ApJS, 223, 26
Ackermann M. et al., 2011, ApJ, 726, 81
Adriani O. et al., 2014, ApJ, 791, 93
Adriani O. et al., 2018, Phys. Rev. Lett., 120, 261102
Aguilar M. et al., 2018, Phys. Rev. Lett., 120, 021101
Aguilar M. et al., 2019a, Phys. Rev. Lett., 122, 041102
Aguilar M. et al., 2019b, Phys. Rev. Lett., 122, 101101
Amato E., Blasi P., 2018, Adv. Space Res., 62, 2731
Armillotta L., Ostriker E. C., Jiang Y.-F., 2021, ApJ, 922, 11
Atkin E. V. et al., 2019, Astron. Rep., 63, 66
Bai X.-N., 2022, ApJ, 928, 112
Bai X.-N., Caprioli D., Sironi L., Spitkovsky A., 2015, ApJ, 809, 55
Bai X.-N., Ostriker E. C., Plotnikov I., Stone J. M., 2019, ApJ, 876, 60
Beattie J. R., Krumholz M. R., Federrath C., Sampson M., Crocker R. M., 2022, Frontiers in Astronomy and Space Sciences, 9, 900900
Bell A. R., 2004, MNRAS, 353, 550
Benincasa S. M. et al., 2020, MNRAS, 497, 3993
Beresnyak A., 2019, Living Rev. Comput. Astrophys., 5, 2
Bindi V., Corti C., Consolandi C., Hoffman J., Whitman K., 2017, Adv. Space Res., 60, 865
Bisschoff D., Potgieter M. S., Aslam O. P. M., 2019, ApJ, 878, 59
Blasi P., Amato E., 2012, J. Cosmol. Astropart. Phys., 1, 010
Boezio M. et al., 2000, ApJ, 532, 653
Boezio M. et al., 2017, in Kwak Y. S., Lee H. S., Oh S., Heung Park I., eds, International Cosmic Ray Conference Vol. 301, 35th International Cosmic Ray Conference (ICRC2017). Proceedings of Science, Sissa Medialab Srl., Italy, p. 1091
Boldyrev S., 2005, ApJ, 626, L37
Boldyrev S., 2006, Phys. Rev. Lett., 96, 115002
Braginskii S. I., 1965, Rev. Plasma Phys., 1, 205
Brandenburg A., Nordlund Å., 2011, Rep. Prog. Phys., 74, 046901
Buck T., Pfrommer C., Pakmor R., Grand R. J. J., Springel V., 2020, MNRAS, 497, 1712
Burgers J., 1973, The nonlinear diffusion equation: asymptotic solutions and statistical problems. D. Reidel Pub. Co., Springer Netherlands, https://link.springer.com/book/10.1007/978-94-010-1745-9#bibliographic-information
Bustard C., Zweibel E. G., 2021, ApJ, 913, 106
Butsky I. S., Quinn T. R., 2018, ApJ, 868, 108
Butsky I. S., Fielding D. B., Hayward C. C., Hummels C. B., Quinn T. R., Werk J. K., 2020, ApJ, 903, 77
Cesarsky C. J., 1971, in Bulletin of the American Astronomical Society. American Astronomical Society Publishing, Washington DC, p. 480
Cesarsky C. J., Kulsrud R. M., 1981, in Setti G., Spada G., Wolfendale A. W., eds, IAU Symp. No. 94 Vol. 94, Origin of Cosmic Rays. D. Reidel Publishing Co, Dordrecht, Dordrecht, p. 251
Chan T. K., Kereš D., Hopkins P. F., Quataert E., Su K. Y., Hayward C. C., Faucher-Giguère C. A., 2019, MNRAS, 488, 3716
Chan T. K., Keres D., Gurvich A. B., Hopkins P., Trapp C., Ji S., Faucher-Giguere C.-A., 2021, MNRAS, 517, 597
Chandran B. D. G., 2000, Phys. Rev. Lett., 85, 4656
Chen C. H. K., 2016, J. Plasma Phys., 82, 535820602
Cho J., Lazarian A., 2003, MNRAS, 345, 325
Colbrook M. J., Ma X., Hopkins P. F., Squire J., 2017, MNRAS, 467, 2421
Cummings A. C. et al., 2016, ApJ, 831, 18
DAMPE Collaboration, 2017, Nature, 552, 63
De La Torre Luque P., Mazziotta M. N., Loparco F., Gargano F., Serini D., 2021, J. Cosmol. Astropart. Phys., 2021, 099
Digel S. W., Grenier I. A., Hunter S. D., Dame T. M., Thaddeus P., 2001, ApJ, 555, 12
Elmegreen B. G., 2002, ApJ, 577, 206
Elsasser K., Schamel H., 1974, Physics Lett. A, 47, 419
Elsässer K., Schamel H., 1976, Z. Phys. B Condens. Matter, 23, 89
Emami N., Siana B., Weisz D. R., Johnson B. D., Ma X., El-Badry K., 2019, ApJ, 881, 71
Engelmann J. J. et al., 1990, A&A, 233, 96
Enßlin T., Pfrommer C., Miniati F., Subramanian K., 2011, A&A, 527, A99
Erlebacher G., Hussaini M. Y., Kreiss H. O., Sarkar S., 1990, Theor. Comput. Fluid Dyn., 2, 73
Escala I. et al., 2018, MNRAS, 474, 2194
Evoli C., Gaggero D., Vittino A., Di Bernardo G., Di Mauro M., Ligorini A., Ullio P., Grasso D., 2017, J. Cosmol. Astropart. Phys., 2, 015
Farber R., Ruszkowski M., Yang H.-Y. K., Zweibel E. G., 2018, ApJ, 856, 112
Farmer A. J., Goldreich P., 2004, ApJ, 604, 671
Faucher-Giguère C.-A., 2020, MNRAS, 493, 1614
Ferrand R., Galtier S., Sahraoui F., Federrath C., 2020, ApJ, 904, 160
Foote E. A., Kulsrud R. M., 1979, ApJ, 233, 302
Fornieri O., Gaggero D., Cerri S. S., De La Torre Luque P., Gabici S., 2021, MNRAS, 502, 5821
Fu L., Xia Z. Q., Shen Z. Q., 2017, MNRAS, 471, 1737
Gaches B. A. L., Offner S. S. R., 2018, ApJ, 861, 87
Gaggero D., Urbano A., Valli M., Ullio P., 2015, Phys. Rev. D, 91, 083012
Galtier S., Banerjee S., 2011, Phys. Rev. Lett., 107, 134501
Galtier S., Nazarenko S. V., Newell A. C., Pouquet A., 2000, J. Plasma Phys., 63, 447
Garrison-Kimmel S. et al., 2019, MNRAS, 487, 1380
Ginzburg V. L., Syrovatsky S. I., 1961, Prog. Theor. Phys. Suppl., 20, 1
Girichidis P. et al., 2016, ApJ, 816, L19
Girichidis P., Pfrommer C., Hanasz M., Naab T., 2020, MNRAS, 491, 993
Goldreich P., Sridhar S., 1995, ApJ, 438, 763
Griffin R. D., Dai X., Thompson T. A., 2016, ApJ, 823, L17
Grudić M. Y., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Murray N., Kereš D., 2018, MNRAS, 475, 3511
Guo Y.-Q., Tian Z., Jin C., 2016, ApJ, 819, 54
Gurvich A. B. et al., 2020, MNRAS, 498, 3664
Guszejnov D., Grudić M. Y., Offner S. S. R., Boylan-Kolchin M., Faucher-Giguère C.-A., Wetzel A., Benincasa S. M., Loebman S., 2020, MNRAS, 492, 488
H. E. S. S. Collaboration, 2017, preprint (arXiv:1709.06442)
Hafen Z. et al., 2019, MNRAS, 488, 1248
Hin Navin Tsung T., Oh S. P., Jiang Y.-F., 2022, MNRAS, 513, 4464
Holcomb C., Spitkovsky A., 2019, ApJ, 882, 3
Holman G. D., Ionson J. A., Scott J. S., 1979, ApJ, 228, 576
Hopkins P. F., 2013, MNRAS, 430, 1880
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., 2016, MNRAS, 462, 576
Hopkins P. F., 2017, MNRAS, 466, 3387
Hopkins P. F., Raives M. J., 2016, MNRAS, 455, 51
Hopkins P. F., Squire J., 2018, MNRAS, 479, 4681
Hopkins P. F., Narayanan D., Murray N., 2013, MNRAS, 432, 2647
Hopkins P. F. et al., 2018a, MNRAS, 477, 1578
Hopkins P. F. et al., 2018b, MNRAS, 480, 800
Hopkins P. F., Grudić M. Y., Wetzel A., Kereš D., Faucher-Giguère C.-A., Ma X., Murray N., Butcher N., 2020a, MNRAS, 491, 3702
Hopkins P. F. et al., 2020b, MNRAS, 492, 3465

Hopkins P. F., Butsky I. S., Panopoulou G. V., Ji S., Quataert E., Faucher-Giguere C.-A., Keres D., 2021a, MNRAS, 516, 3470
Hopkins P. F., Chan T. K., Ji S., Hummels C. B., Kereš D., Quataert E., Faucher-Giguère C.-A., 2021b, MNRAS, 501, 3640
Hopkins P. F., Chan T. K., Squire J., Quataert E., Ji S., Kereš D., Faucher-Giguère C.-A., 2021c, MNRAS, 501, 3663
Hopkins P. F., Squire J., Chan T. K., Quataert E., Ji S., Kereš D., Faucher-Giguère C.-A., 2021d, MNRAS, 501, 4184
Hopkins P. F. et al., 2022a, MNRAS, submitted, preprint (arXiv:2203.00040)
Hopkins P. F., Squire J., Butsky I. S., 2022b, MNRAS, 509, 3779
Huang X., Davis S. W., 2022, MNRAS, 511, 5125
Indriolo N., McCall B. J., 2012, ApJ, 745, 91
Indriolo N., Fields B. D., McCall B. J., 2009, ApJ, 694, 257
Indriolo N. et al., 2015, ApJ, 800, 40
Iroshnikov P. S., 1963, Astron. Zh., 40, 742
Isenberg P. A., 1997, J. Geophys. Res., 102, 4719
Ji S. et al., 2020, MNRAS, 496, 4221
Ji S., Kereš D., Chan T. K., Stern J., Hummels C. B., Hopkins P. F., Quataert E., Faucher-Giguère C.-A., 2021, MNRAS, 505, 259
Jóhannesson G. et al., 2016, ApJ, 824, 16
Jokipii J. R., 1966, ApJ, 146, 480
Kachelrieß M., Semikoz D. V., 2019, Prog. Part. Nucl. Phys., 109, 103710
Kadomtsev B. B., Petviashvili V. I., 1973, Sov. Phys. Dokl., 18, 115
Keating L. C. et al., 2020, MNRAS, 499, 837
Kempski P., Quataert E., 2022, MNRAS, 514, 657
Kim J. et al., 2022, ApJ, 926, 179
Kolmogorov A., 1941, Akad. Nauk SSSR Dokl., 30, 301
Konstandin L., Federrath C., Klessen R. S., Schmidt W., 2012, J. Fluid Mech., 692, 183
Korsmeier M., Cuoco A., 2016, Phys. Rev. D, 94, 123019
Kowal G., Lazarian A., 2010, ApJ, 720, 742
Kraichnan R. H., 1965, Phys. Fluids, 8, 1385
Kulsrud R., Pearce W. P., 1969, ApJ, 156, 445
Kuznetsov E., Krasnoselskikh V., 2008, Phys. Plasmas, 15, 062305
Lacki B. C., Thompson T. A., Quataert E., Loeb A., Waxman E., 2011, ApJ, 734, 107
Lazarian A., 2016, ApJ, 833, 131
Le Roux J. A., Matthaeus W. H., Zank G. P., 2001, Geophys. Res. Lett., 28, 3831
Le Roux J. A., Zank G. P., Li G., Webb G. M., 2005, ApJ, 626, 1116
Le Roux J. A., Zank G. P., Webb G. M., Khabarova O., 2015, ApJ, 801, 112
Lee E. J., Hopkins P. F., 2020, MNRAS, 495, L86
Lee M. A., Völk H. J., 1973, Ap&SS, 24, 31
Lee E., Brachet M. E., Pouquet A., Mininni P. D., Rosenberg D., 2010, Phys. Rev. E, 81, 016318
Lopez L. A., Auchettl K., Linden T., Bolatto A. D., Thompson T. A., Ramirez-Ruiz E., 2018, ApJ, 867, 44
Mac Low M.-M., Klessen R. S., 2004, Rev. Mod. Phys., 76, 125
Makwana K. D., Yan H., 2020, Phys. Rev. X, 10, 031021
Maurin D., 2020, Comput. Phys. Commun., 247, 106942
Mee A. J., Brandenburg A., 2006, MNRAS, 370, 415
Obermeier A., Ave M., Boyle P., Höppner C., Hörandel J., Müller D., 2011, ApJ, 742, 14
Padovani M., Galli D., Glassgold A. E., 2009, A&A, 501, 619
Parker R. J., 2020, R. Soc. Open Sci., 7, 201271
Ponnada S. B. et al., 2022, MNRAS, 516, 4417
Quataert E., Thompson T. A., Jiang Y.-F., 2022a, MNRAS, 510, 1184
Quataert E., Jiang Y.-F., Thompson T. A., 2022b, MNRAS, 510, 920
Ruszkowski M., Yang H.-Y. K., Zweibel E., 2017, ApJ, 834, 208
Salem M., Bryan G. L., 2014, MNRAS, 437, 3312
Salem M., Bryan G. L., Corlies L., 2016, MNRAS, 456, 582
Samuel J. et al., 2020, MNRAS, 491, 1471
Scalo J., Elmegreen B. G., 2004, ARA&A, 42, 275
Schekochihin A. A., 2022, Journal of Plasma Physics, 88, 155880501
Schekochihin A. A., Cowley S. C., Dorland W., Hammett G. W., Howes G. G., Quataert E., Tatsuno T., 2009, ApJS, 182, 310
Schlickeiser R., 1989, ApJ, 336, 243
Schmidt W., Federrath C., Hupp M., Kern S., Niemeyer J. C., 2009, A&A, 494, 127
Shikaze Y. et al., 2007, Astropart. Phys., 28, 154
Shivamoggi B. K., 1992, Phys. Lett. A, 166, 243
Shivamoggi B. K., 2011, Phys. A Stat. Mech. Appl., 390, 1534
Simpson C. M., Pakmor R., Marinacci F., Pfrommer C., Springel V., Glover S. C. O., Clark P. C., Smith R. J., 2016, ApJ, 827, L29
Skilling J., 1971, ApJ, 170, 265
Skilling J., 1975a, MNRAS, 172, 557
Skilling J., 1975b, MNRAS, 173, 255
Spitzer L., 1978, Physical Processes in the Interstellar Medium. Wiley-Interscience, New York
Squire J., Hopkins P. F., 2017, MNRAS, 471, 3753
Squire J., Hopkins P. F., 2018, ApJ, 856, L15
Squire J., Quataert E., Schekochihin A. A., 2016, ApJ, 830, L25
Squire J., Schekochihin A. A., Quataert E., 2017, New J. Phys., 19, 055005
Squire J., Hopkins P. F., Quataert E., Kempski P., 2021, MNRAS, 502, 2630
Sridhar S., Goldreich P., 1994, ApJ, 432, 612
Strong A. W., Moskalenko I. V., 2001, Adv. Space Res., 27, 717
Strong A. W., Moskalenko I. V., Ptuskin V. S., 2007, Annu. Rev. Nucl. Part. Sci., 57, 285
Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, MNRAS, 471, 144
Su K.-Y., Hayward C. C., Hopkins P. F., Quataert E., Faucher-Giguère C.-A., Kereš D., 2018, MNRAS, 473, L111
Su K.-Y. et al., 2020, MNRAS, 491, 1190
Su K.-Y. et al., 2021, MNRAS, 507, 175
Sun B., 2016, Mod. Phys. Lett. B, 30, 1650297
Tang Q.-W., Wang X.-Y., Tam P.-H. T., 2014, ApJ, 794, 26
Terry P. W., 2018, Phys. Plasmas, 25, 092301
Thomas T., Pfrommer C., 2019, MNRAS, 485, 2977
Thomas T., Pfrommer C., Pakmor R., 2022, preprint (arXiv:2203.12029)
Thompson T. A., 2013, MNRAS, 431, 63
Tibaldo L., 2014, Braz. J. Phys., 44, 600
Tibaldo L. et al., 2015, ApJ, 807, 161
Tibaldo L., Gaggero D., Martin P., 2021, Universe, 7, 141
Trapp C. W. et al., 2022, MNRAS, 509, 4149
Tytarenko P. V., Williams R. J. R., Falle S. A. E. G., 2002, MNRAS, 337, 117
Van Marle A. J., Casse F., Marcowith A., 2019, MNRAS, 490, 1156
Vladimirov A. E., Jóhannesson G., Moskalenko I. V., Porter T. A., 2012, ApJ, 752, 68
Voelk H. J., 1975, Rev. Geophys. Space Phys., 13, 547
Völk H. J., Cesarsky C. J., 1982, Z. Nat.forsch A, 37, 809
Volk H. J., McKenzie J. F., 1981, in ICRC ed., International Cosmic Ray Conference Vol. 9, Proc. 17th International Cosmic Ray Conference. Gif-sur-Yvette, Essonne, France, Commissariat a l'Energie Atomique, Gif-sur-Yvette Essonne, France, p. 246
Wang X., Fields B. D., 2018, MNRAS, 474, 4073
Weingartner J. C., Draine B. T., 2001, ApJ, 553, 581
Wellons S. et al., 2022, preprint (arXiv:2203.06201)
Wentzel D. G., 1968, ApJ, 152, 987
Wentzel D. G., 1969, ApJ, 156, 303
Wiener J., Oh S. P., Guo F., 2013a, MNRAS, 434, 2209
Wiener J., Zweibel E. G., Oh S. P., 2013b, ApJ, 767, 87
Wiener J., Pfrommer C., Oh S. P., 2017, MNRAS, 467, 906
Wojaczyński R., Niedźwiecki A., 2017, ApJ, 849, 97
Wolfire M. G., Hollenbach D., McKee C. F., Tielens A. G. G. M., Bakes E. L. O., 1995, ApJ, 443, 152
Yan H., Lazarian A., 2002, Phys. Rev. Lett., 89, 281102
Yan H., Lazarian A., 2004, ApJ, 614, 757
Yan H., Lazarian A., 2008, ApJ, 673, 942
Yang R., Aharonian F., Evoli C., 2016, Phys. Rev. D, 93, 123007
Yoon Y. S. et al., 2017, ApJ, 839, 5
Zakharov V. E., Sagdeev R. Z., 1970, Sov. Phys. Dokl., 15, 439
Zank G. P., 2014, Transport Processes in Space Physics and Astrophysics. Lecture Notes in Physics Vol. 877. Springer Science + Business Media, New York
Zweibel E. G., 2013, Phys. Plasmas, 20, 055501

Zweibel E. G., 2017, Phys. Plasmas, 24, 055402

APPENDIX A: EQUILIBRIUM SELF-CONFINEMENT MODELS

A1 Basic equations and setup

Here, we consider the behaviour of SC models in steady-state. Given that (as shown in Section B below) the CR flux $F'_{e,\rm cr}$ and $e_\pm$ equations converge to local steady state on a time-scale much shorter than the CR energy $e'_{\rm cr}$ equation, we can safely assume their steady-state values (in Section B) in evaluating the CR energy equation. We will assume only SC driving (take $S_{\rm ext,\pm}=S_{\rm new,\pm}=0$), giving [from equation (11)]:

$$D_t e'_{\rm cr} \rightarrow -\nabla\cdot\left(F'_{e,\rm cr}\,\hat{\mathbf{b}}\right) + \mathcal{S}'_{\rm eff} \tag{A1}$$

where $F'_{e,\rm cr}\equiv -(v_{\rm cr}^2/3\,\bar{\nu}_{\rm s})\,\hat{\mathbf{b}}\cdot\nabla e'_{\rm cr} + \bar{v}_A\,e'_{\rm cr}$ includes the 'diffusive' ($\propto\bar{\nu}_{\rm s}^{-1}$) and 'streaming' ($\propto\bar{v}_A$, with $\bar{v}_A\rightarrow -v_{A,\rm eff}\,{\rm sign}[\hat{\mathbf{b}}\cdot\nabla e'_{\rm cr}]$) terms, and $S'_{\rm eff}\equiv -P'_{\rm cr}\,\nabla\cdot(\mathbf{u}_{\rm gas}+\bar{v}_A\,\hat{\mathbf{b}}) + \mathcal{S}'_{\rm other,\,cr}$ includes the 'adiabatic' and 'streaming loss' terms ($\propto\nabla\cdot[\mathbf{u}_{\rm gas}+\bar{v}_A\,\hat{\mathbf{b}}]$) and all injection and radiative/catastrophic losses in $\mathcal{S}'_{\rm other,\,cr}$.

In steady-state ($D_t e'_{\rm cr}\rightarrow 0$), integrating equation (A1) over some volume V with surface ∂V immediately gives:

$$\left\langle F'_{e,\rm cr}\right\rangle\,A_{\rm eff}\equiv\oint_{\partial V}F'_{e,\rm cr}\hat{\mathbf{b}}\cdot{\rm d}\mathbf{A}=\int_V{\rm d}^3\mathbf{x}\,\mathcal{S}'_{\rm eff}\equiv\dot{E}_{\rm inj,\,eff}. \tag{A2}$$

Here, $\langle F'_{e,\rm cr}\rangle$ is the weighted-mean scalar flux from the integral over $F'_{e,\rm cr}$, $A_{\rm eff}\equiv\oint_{\partial V}|{\rm d}\mathbf{A}|$, and $\dot{E}_{\rm inj,\,eff}$ is the net CR energy production inside ∂V. For simplicity, we will consider CR primary species at rigidities $\gtrsim$GV where $v_{\rm cr}\sim c$ (so $P'_{\rm cr}\approx e'_{\rm cr}/3$) and empirical constraints (see text and Paper I) indicate losses are negligible, so $\dot{E}_{\rm inj,\,eff}\approx\int_V{\rm d}^3\mathbf{x}\,j_{{\rm inj},\,e}(R_{\rm cr})=\dot{E}_{\rm inj}$ is approximately the total injection rate.

A2 Behaviour of phenomenological or ET models

First, consider a typical phenomenological model, where $\bar{\nu}_{\rm s}$ is taken to be constant with $\bar{\nu}_{\rm s}\sim 10^{-9}\,{\rm s}^{-1}\,R_{\rm GV}^{-0.5}$ as in Fig. 1. With $\bar{\nu}_{\rm s}=$constant, equation (A1) indeed behaves like a diffusion equation, with the diffusive term much larger than streaming terms on scales of interest, and if we assume tangled magnetic fields, the effective isotropic flux is just $\langle F'_{e,\rm cr}\rangle\approx\kappa_{\rm iso}\,\langle|\nabla e'_{\rm cr}|\rangle$ with $\kappa_{\rm iso}\equiv(c^2/9\,\bar{\nu}_{\rm s})$. For the Galaxy, take $\dot{E}_{\rm inj,\,eff}\approx\dot{E}_{\rm inj}\approx 0.1\,\dot{E}_{\rm SNe}\,f_{\rm inj}(R_{\rm cr})\sim 3\,\dot{N}_{\rm SNe,\,100}\,R_{\rm GV}^{-0.2}\times 10^{40}\,{\rm erg\,s^{-1}}$ where $\dot{N}_{\rm SNe,\,100}$ is the SNe rate inside ∂V in units of $\sim 1/(100\,{\rm yr})$ and $f_{\rm inj}(R_{\rm cr})\equiv(1/\dot{E}_{\rm inj}^{\rm total})\,{\rm d}\dot{E}_{\rm inj}^{\rm total}/{\rm d}\ln R_{\rm cr}\sim R_{\rm GV}^{-0.2}$ is the fraction injected at the given $R_{\rm cr}$ (according to our assumed standard injection slope in the text). Assuming e.g. spherical symmetry or a vertically stratified model, the steady-state $e'_{\rm cr}$ profile is then trivially solved by $\nabla e'_{\rm cr}=-\dot{E}_{\rm in}/(\kappa_{\rm iso}\,A_{\rm eff})$. If we assume approximate spherical symmetry at large Galactocentric radii $\langle F'_{e,\rm cr}\rangle=\dot{E}_{\rm inj}/(4\pi\,r^2)$, we obtain $e'_{\rm cr}\sim 0.6\,{\rm eV\,cm^{-3}}\,\dot{N}_{\rm SNe,\,100}\,R_{\rm GV}^{-0.7}$ at the Solar circle ($r\sim 8.3$ kpc), in excellent agreement with observations (by construction, of course, since $\bar{\nu}_{\rm s}$ was originally fit to the data).

In standard ET models, $\bar{\nu}_s\rightarrow\bar{\nu}_s(k_\parallel,\,\mathbf{B},\,\ell_A,\,f_{\rm ion},\,\ldots)$ can be some arbitrary function of ISM properties, but (crucially) is – like in the phenomenological model above – *independent* of $e'_{\rm cr}$ ($\zeta_{\rm cr}=\xi_{\rm cr}=0$). This means that, again, solutions always exist for a steady-state $e'_{\rm cr}$ profile, given by solving equation (A2): $\nabla e'_{\rm cr}\sim-\dot{E}_{\rm in}/(\kappa_{\rm iso}\,A_{\rm eff})\sim -9\,\dot{E}_{\rm in}\,\bar{\nu}_{\rm s}/(c^2\,A_{\rm eff})=\mathcal{F}(r,\,k_\parallel,\,\mathbf{B},\,\ell_A,\,f_{\rm ion},\,\ldots)$. Whether or not these solutions have the correct observed behaviour (as a function of e.g. CR rigidity) is what we investigate in the main text.

A3 Behaviour of SC models

But now consider SC models, with $\bar{\nu}_{\rm s}\approx(3\pi\,\Omega_{\rm cr}/16)\,(e_A/e_{\rm B})$ where e_A is set (see Section B) by the competition between damping ($\Gamma_\pm$) and driving with $S_{\rm sc}\rightarrow|v_{A,\rm eff}\,\hat{\mathbf{b}}\cdot\nabla P'_{\rm cr}|$, giving $e_A/e_{\rm B}\rightarrow(\Gamma_{\rm lin}/2\,\Gamma^0_{\rm nll})\,(-1+[1+4\,S_{\rm sc}\,\Gamma^0_{\rm nl}/\,\Gamma^2_{\rm lin}\,e_{\rm B}]^{1/2})$, where $\Gamma_{\rm lin}\equiv\Gamma_{\rm in}+\Gamma_{\rm turb/LL}+\Gamma_{\rm dust}+\Gamma_{\rm new,\,damp}+\ldots$ collects all linear damping terms and $\Gamma^0_{\rm nl}$ collects the pre-factors of any non-linear terms (e.g. $\Gamma^0_{\rm nl}=\Gamma^0_{\rm nll}=(\sqrt{\pi}/8)\,c_s\,k_\parallel$ for NLL damping). The dependence of e_A on $e'_{\rm cr}$ introduces fundamentally distinct behaviour.

A3.1 Linear damping

First, assume that linear damping dominates[20] ($\Gamma_{\rm lin}\gtrsim\Gamma_{\rm nl}=\Gamma^0_{\rm nl}\,(e_A/e_{\rm B})$). Then, $e_A\rightarrow S_{\rm sc}/\Gamma_{\rm lin}$, giving the 'diffusive' flux $F'_{e,\rm cr}\rightarrow(16/3\pi)\,(c\,e_{\rm B}\,r_{g,\rm cr}\,\Gamma_{\rm lin})/v_{A,\rm eff}=(4c/3\pi^{3/2})\,R_{\rm cr}\,\rho_{\rm ion}^{1/2}\,\Gamma_{\rm lin}$. If $D_t e'_{\rm cr}$ is dominated by the diffusive term, then $\Gamma_{\rm lin}$ for any known damping mechanism depends on properties extrinsic to the CRs (e.g. turbulent velocities, $e_{\rm B}$, etc.) and so, it and therefore $F'_{e,\rm cr}$ are independent of the CR energy density, this means there exist *no steady-state solutions* for $e'_{\rm cr}$. It does not seem possible to construct an $e'_{\rm cr}$ profile that ensures $\langle F'_{e,\rm cr}\rangle=\dot{E}_{\rm inj,\,eff}/A_{\rm eff}$.[21]

In practice, what this means is that there are only two real equilibrium solutions: if $\langle F'_{e,\rm cr}\rangle<\dot{E}_{\rm inj}/A_{\rm eff}$, since the diffusive flux is independent of $e'_{\rm cr}$, the CR energy density will continue to build up (increasing $\bar{\nu}_{\rm s}\propto e'_{\rm cr}$ and lowering the effective diffusivity or streaming speed) until the streaming term $\propto v_{A,\rm eff}\,e'_{\rm cr}$ dominates $F'_{e,\rm cr}$ or catastrophic loss terms (also $\propto e'_{\rm cr}$) dominate $D_t e'_{\rm cr}$. Thus, CRs collapse to the Alfvénic streaming and/or calorimetric limit, with maximal isotropically averaged streaming speed $\approx v_{A,\rm eff}/2$. This is problematic for two reasons: first, the implied residence time (neglecting losses) to escape the Galaxy and CR scattering halo (~ 10 kpc) is $\sim 10\,{\rm Gyr}\,n_1^{1/2}\,B_{\mu\rm G}^{-1}$, far longer than observationally allowed. Secondly, even if we arbitrarily renormalized the Alfvén speed and/or Galaxy + halo size, the streaming/escape/residence time would (by definition) be *independent* of CR energy (i.e. $\delta_{\rm s}=0$), also ruled out. Alternatively, if $\langle F'_{e,\rm cr}\rangle>\dot{E}_{\rm inj}/A_{\rm eff}$, then $e'_{\rm cr}$ will deplete until $\bar{\nu}_{\rm s}$ is so low[22] that the CRs free-stream and escape at $\sim c$, vastly faster than observed (residence times $\lesssim 10^{4.5}$ yr), with again $\delta_{\rm s}=0$.

An alternative way to see this is to simply insert the full expression for the SC-predicted $\bar{\nu}_{\rm s}$ directly into equation (A1). As noted by many going back to Skilling (1971) and Cesarsky (1971), the 'diffusive' part of the flux-gradient term then formally takes the form of a source

[20] So as long as $\Gamma_{\rm lin}\gtrsim\Gamma_{\rm nl}$, explicitly including small-but-finite $\Gamma_{\rm nl}$ changes none of our conclusions above. The limit $\Gamma_{\rm lin}\lesssim\Gamma_{\rm nl}$ is discussed below.

[21] One might imagine a (contrived) special case where the properties (e.g. $\mathbf{B}$, c_s) which enter $\Gamma_{\rm lin}$ scale exactly as required with both $\dot{E}_{\rm inj}$ and position $\mathbf{x}$ such that $\langle F'_{e,\rm cr}\rangle=\dot{E}_{\rm inj}/A_{\rm eff}$. However, not only does this require exceptional fine-tuning, but (1) because $\dot{E}_{\rm inj}$ and $F'_{e,\rm cr}$ scale *differently* with $R_{\rm GV}$, it is impossible to satisfy this at any two CR energies simultaneously, and (2) such a solution cannot 'respond' to adjust to any perturbations to the source rate $\dot{E}_{\rm inj}$ or to the gas quantities which enter $F'_{e,\rm cr}$.

[22] As noted in the main text, if ET is present, then at some sufficiently low $\bar{\nu}_{\rm s}$, $S_{\rm et,\pm}$ will dominate so $\bar{\nu}_{\rm s}$ will not vanish entirely, but in this case the system is in the entirely ET-dominated limit.

or sink term:

$$D_t e'_{\rm cr} = \pm\left(\frac{4\,c\,R_{\rm cr}}{3\,\pi^{3/2}}\right)\,\nabla\cdot(\rho_{\rm ion}^{1/2}\,\Gamma_{\rm lin}\,\hat{\mathbf{b}}) \sim \pm\frac{{\rm eV\,cm^{-3}}}{\rm Myr}\,B_{\rm \mu G}^{3/2}\,R_{\rm GV}^{1/2}\,\left(\frac{10\,{\rm pc}}{\ell_{\nabla,\Gamma\rho^{1/2}}}\right) \quad \text{(A3)}$$

where the sign is determined by the gradient in $e'_{\rm cr}$; $\ell_{\nabla,\Gamma\rho^{1/2}} \equiv \rho_{\rm ion}^{1/2}\,\Gamma_{\rm lin}/|\nabla\cdot(\rho_{\rm ion}^{1/2}\,\Gamma_{\rm lin}\,\hat{\mathbf{b}})|$ is the gradient scale length of $f_{\rm ion}^{1/2}\,\rho^{1/2}\,\Gamma_{\rm lin}$, which can vary on $\lesssim$pc scales (Hopkins et al. 2021d); and in the second equality, we inserted the scalings for $\Gamma_{\rm lin}$ for turbulent/linear Landau damping (to give a typical value). From equation (A3), it is clear that within a time-scale $\sim$Myr, the CR energy $e'_{\rm cr}$ will either (1) be driven to negligible values if $D_t e'_{\rm cr}$ is negative (making all other terms in $D_t e'_{\rm cr}$ smaller, until the 'free escape' limit is reached), or (2) be driven to increase if $D_t e'_{\rm cr}$ is positive, until the other terms in $D_t e'_{\rm cr}$ such as the streaming term $\propto \bar{v}_{A,\,\rm eff}\,e'_{\rm cr}$ dominate (the 'over-confined' limit).

A3.2 Non-linear damping

Now, instead, assume NLL damping dominates. Let us first ask when this might occur: for NLL damping to set e_A (see Section B) requires the dimensionless $\psi_{\rm nl} \equiv |4\,S_{\rm sc}\,\Gamma_{\rm nl}^0/\Gamma_{\rm lin}^2\,e_{\rm B}|^{1/2} \gg 1$. Taking the standard linear damping scalings from Section 2.5, $\psi_{\rm nl} \gg 1$ requires the gas is highly ionized ($f_{\rm neutral} \lesssim 10^{-3}$, so $\Gamma_{\rm in}$ is small), has a low dust-to-gas ratio ($f_{\rm dg} \lesssim 10^{-3}$, so $\Gamma_{\rm dust}$ is small), and is weakly turbulent with a high CR energy density at the given $R_{\rm cr}$ ($\mathcal{M}_A^2 \lesssim e'_{\rm cr}/{\rm eV\,cm^{-3}}$, so $\Gamma_{\rm turb/LL} \ll \Gamma_{\rm nll}$). While specific, this is not impossible around $\sim$ 1 GV where $e'_{\rm cr}$ peaks, in the diffuse warm/hot ISM/CGM.

Assuming $\psi_{\rm nl} \gg 1$ with NLL damping dominating, we have $e_{\rm A} \rightarrow (S_{\rm sc}\,e_{\rm B}/\Gamma_{\rm nl}^0)^{1/2}$, so the 'diffusive' $F'_{e,\,\rm cr} \rightarrow (2^{5/2}/3^{3/2}\,\pi^{3/4})\,c\,(|\hat{\mathbf{b}}\cdot\nabla e'_{\rm cr}|\,c_s\,e_{\rm B}\,r_{g,\rm cr}/v_A)^{1/2}$. This does formally have a steady-state solution given by $\langle|\hat{\mathbf{b}}\cdot\nabla e'_{\rm cr}|\rangle \approx (27\pi^{3/2}/32)\,(v_{A,\,\rm eff}/c^2\,c_s\,e_{\rm B}\,r_{g,\rm cr})\,(\dot{E}_{\rm inj}/A_{\rm eff})^2$ with $\langle\bar{\nu}_{\rm s}\rangle \rightarrow (9\pi^{3/2}/32)\,(v_{A,\,\rm eff}\,\dot{E}_{\rm inj}/c_s\,e_{\rm B}\,r_{g,\rm cr}\,A_{\rm eff})$. But this solution has some very strange features: using $A_{\rm eff} \sim 4\pi\,r^2$ as above and evaluating it at the Solar circle we obtain: $e'_{\rm cr} \approx 2000\,{\rm eV\,cm^{-3}}\,R_{\rm GV}^{-1.4}\,N_{\rm SNe,\,100}^2\,(n_1\,T_4)^{-1/2}$ (with $T_4 = T/10^4\,{\rm K}$) and $\bar{\nu}_{\rm s} \rightarrow 10^{-5}\,{\rm s^{-1}}\,R_{\rm GV}^{-1.2}\,N_{\rm SNe,\,100}\,(n_1\,T_4)^{-1/2}$. These are enormously unphysically high-CR energy densities and scattering rates, which also exhibit a clearly ruled-out dependence on $R_{\rm GV}$. In practice, this means that, beginning from any physically realistic (much smaller) $e'_{\rm cr}$, $F'_{e,\,\rm cr} \ll \dot{E}_{\rm cr}/A_{\rm eff}$ will drive $D_t e'_{\rm cr} > 0$ so $e'_{\rm cr}$ increases until either the streaming or loss terms (which scale $\propto e'_{\rm cr}$, while the diffusive term scales $\propto \sqrt{e'_{\rm cr}}$) dominate $D_t e'_{\rm cr}$ (e.g. the streaming flux will dominate $F'_{e,\,\rm cr}$ once $e'_{\rm cr} \gtrsim 0.5\,{\rm eV\,cm^{-3}}\,R_{\rm GV}\,(n_1^{3/2}\,T_4^{1/2}\,B_{\rm \mu G}^{-2}\,{\rm kpc}/\ell_{\nabla,\rm cr})$).[23]

So again, we see immediate 'solution collapse,' but the conditions where non-linear damping dominates, which require higher e_A and therefore higher $e'_{\rm cr}$, are such that they always drive the collapse to the over-confined/streaming solution.

[23]From the scalings above, in this limit the flux should be dominated by the Alfvénic streaming component at all galacto-centric radii interior to $\lesssim$ Mpc $B_{\rm \mu G}\,N_{\rm SNe,\,100}\,n_1^{-1}\,T_4^{-1/2}\,R_{\rm GV}^{-1.2}$.

A3.3 Summary

These behaviours above are what we refer to in the text as the SC models being globally 'not stable.' This is not necessarily a linear-stability analysis (in fact the 'collapsed' free streaming or over-confined spherical equilibrium solutions above, if we perturb just $e'_{\rm cr}$ infinitesimally and ignore all interactions with the gas, are formally linearly stable). But it is common, when SC is discussed, to refer to 'super-Alfvénic streaming,' i.e. flux in excess of $v_{A,\,\rm eff}\,e'_{\rm cr}$ with an effective contribution to $F'_{e,\,\rm cr}$ from the $\bar{\nu}_{\rm s}$ term as defined above (i.e. a finite-but-not-infinite CR transport speed in excess of $v_{A,\,\rm eff}$). This can arise trivially, in any infinitesimal local patch, if one defines $\bar{\nu}_{\rm s}$ for a *given* $e'_{\rm cr}$ (e.g. choosing a fixed $e'_{\rm cr}$ similar to the Solar circle value) – in fact we show this below in Section B, where we derive the values of $e_\pm$ given by assuming local steady-state of the CR flux equation. But generically, these solutions will *not* give a self-consistent steady-state for the CR *energy* equation: converging to such a 'local equilibrium' value of $\bar{\nu}_{\rm s}$ for a given $e'_{\rm cr}$ (as determined by the CR flux and $e_\pm$ equations, which evolve on very short time-scales) will mean necessarily that the energy equation is out-of-steady-state. This then forces $e'_{\rm cr}$, and correspondingly $\bar{\nu}_{\rm s}$, to evolve either towards 'bottleneck' and the infinite-strong-scattering Alfvénic streaming regime, or towards 'escape,' de-confinement, and the negligible-scattering streaming-at-c regime. Any ICs rapidly collapses (over $\sim$Myr) towards one of these two states for all spatial and CR energy scales of interest if $S_{\rm sc}$ is the dominant driving term.

APPENDIX B: LOCAL STEADY-STATE SOLUTIONS FOR SCATTERING RATES IN DETAIL

B1 Relevant equations and limits

Consider the CR flux and $e_\pm$ equations in more detail. From the general versions of equations (3), (7), and (9), in the text, we can write:

$$D_t F'_{e,\,\rm cr} + c^2\,\hat{\mathbf{b}}\cdot\left(\nabla\cdot\mathbb{P}'_{\rm cr}\right) = -\frac{\omega}{e_{\rm B}}\,(e_+ + e_-)\,F'_{e,\,\rm cr} + \frac{\omega}{e_{\rm B}}\,(e_+ - e_-)\,H'_{\rm cr} \quad \text{(B1)}$$

$$D_t e_\pm + \nabla\cdot\left(v_{A,\,\pm}\,e_\pm\,\hat{\mathbf{b}}\right) = -\frac{e_\pm}{2}\nabla\cdot\mathbf{u}_{\rm gas} - \Gamma_{\rm L}\,e_\pm - \Gamma_{\rm NL}\,\frac{e_\pm}{e_{\rm B}}\,e_\pm + \frac{\omega\,v_{A,\,\pm}\,e_\pm}{c^2\,e_{\rm B}}\,\left(F'_{e,\,\rm cr} - H'_{\rm cr}\right) + S_{\rm et,\,\pm} \quad \text{(B2)}$$

where $H'_{\rm cr} \equiv 3\,\chi\,v_{A,\,\pm}\,(e'_{\rm cr} + P'_{\rm cr})$, $\omega \equiv (\pi\,\hat{\nu}_{\rm s}/4)\,\Omega_{\rm cr}$, $v_{A,\,\pm} \equiv \pm v_{A,\,\rm eff}$, and we have expanded the damping rates in terms of the various linear ($\xi_A = 0$, $\Gamma_{\rm L}$) and non-linear ($\xi_A = 1$, $\Gamma_{\rm NL}$) damping terms as $Q_\pm = \Gamma_{\rm L}\,e_\pm + \Gamma_{\rm NL}\,(e_\pm/e_{\rm B})\,e_\pm$ (so for e.g. NLL damping, $\Gamma_{\rm NL} = (\sqrt{\pi}/8)\,c_s\,k_\parallel$).

As noted in the text, these equations evolve towards local steady-state on a time-scale $\sim \bar{\nu}_{\rm s}^{-1} \sim 30\,{\rm yr}\,R_{\rm GV}^{0.5}$ (if we take empirically estimated $\bar{\nu}_{\rm s}$ values), much faster than the time-scales for the CR energy equation or bulk ISM fluid motion time-scales on the scales of interest. Similarly, the 'gradient terms' in equation (B2; the $\nabla\cdot(v_{A,\,\pm}\,e_\pm\,\hat{\mathbf{b}})$ and $\nabla\cdot\mathbf{u}_{\rm gas}$) involve time-scales of order those same ISM time-scales and are much smaller than the other terms in equation (B2). We will justify these assumptions more formally below. Let us therefore assume these equations reach *local* steady-state ($D_t \rightarrow 0$, or $|D_t| \ll \bar{\nu}_s$) in the comoving Alfvén frame [neglecting the 'gradient terms' in equation (B2)] – although we stress that this does not mean the CR *energy* equation is in steady state. In this case, we can re-write them in the dimensionless form:

$$\tilde{g} = -(x_+ + x_-)\,\tilde{f} + (x_+ - x_-)\,\tilde{h}, \tag{B3}$$

$$\pm\,\tilde{f}\,x_\pm = \tilde{h}\,x_\pm + \gamma_L\,x_\pm + \gamma_{NL}\,x_\pm^2 - s_{et}, \tag{B4}$$

where $x_\pm \equiv e_\pm/e_B$, $\tilde{f} \equiv (v_{A,\rm eff}/c)^2\,(F'_{e,\rm cr}/e_B\,v_{A,\rm eff})$, $\tilde{h} \equiv (v_{A,\rm eff}/c)^2\,(H'_{\rm cr}/e_B\,v_{A,\rm eff})$, $\tilde{g} \equiv (v_{A,\rm eff}/\omega\,e_B)\,\hat{\mathbf{b}}\cdot(\nabla\cdot\mathbb{P}'_{\rm cr})$, $\gamma_L \equiv \Gamma_L/\omega$, $\gamma_{NL} \equiv \Gamma_{NL}/\omega$, and $s_{et} \equiv (S_{et,+} + S_{et,-})/(2\,\omega\,e_B)$. This has solutions

$$4\,s_{et} = \bar{x}\,\left[2\,\left(\gamma_L + \tilde{h}\right) + \gamma_{NL}\,\bar{x}\right] - \frac{\tilde{g}^2\,\left[2\,\left(\gamma_L + \tilde{h}\right) + \gamma_{NL}\,\bar{x}\right]}{\bar{x}\,\left(\gamma_L + \gamma_{NL}\,\bar{x}\right)^2} \tag{B5}$$

where $\bar{x} \equiv x_+ + x_-$, with $x_- = x_+ + \tilde{g}/(\gamma_L + \gamma_{NL}\,\bar{x})$ and $\tilde{f} \equiv [(x_+ - x_-)\,\tilde{h} - \tilde{g}]/\bar{x}$ following immediately.

B2 Local steady-state behaviour in ET and SC limits

Unfortunately, equation (B5) is still a fifth-order polynomial for $\bar{x}$, whose general solutions are neither closed-form analytic nor particularly instructive. The solutions do, however, become simple in various limits. First consider the case where linear damping dominates over non-linear (γ_{NL} can be neglected). Then[24] $\bar{x} \rightarrow s_{et}\,(1+\sqrt{1+\Phi^2})/(\gamma_L + \tilde{h})$ with

$$|\Phi| \equiv \frac{|\tilde{g}|}{s_{et}}\,\left[1 + \frac{\tilde{h}}{\gamma_L}\right] \sim \frac{v_{A,\rm eff}\,|\nabla P'_{\rm cr}|}{S_{et,\pm}}\,\left[1 + \frac{\pi}{2}\,\frac{\Omega_{\rm cr}\,e'_{\rm cr}}{\Gamma_L\,\rho_{\rm ion}\,c^2}\right]. \tag{B6}$$

Small $|\Phi| \ll 1$ here corresponds to the ET limit, large $|\Phi| \gg 1$ to the SC limit.

Now consider each of those limits (SC and ET-dominated) in turn, but retain the non-linear term γ_{NL}.

In the ET limit ($|\Phi| \ll 1$): the dimensionless 'streaming speed' $\bar{v}_A/v_{A,\rm eff} = (x_+ - x_-)/(x_+ + x_-) \rightarrow \Phi/2 \ll 1$ is small and $\bar{x} \rightarrow \gamma_{NL}^{-1}\,(\gamma_L + \tilde{h})\,(-1 + [1 + 4\,\gamma_{NL}\,s_{et}/(\gamma_L + \tilde{h})]^{1/2})$, which corresponds to[25] $e_+ \approx e_- \sim S_{et}/(\Gamma_L\,[1+\phi])$ (with $\phi \equiv \tilde{h}/\gamma_L \sim (v_{A,\rm eff}/c)^2\,(\pi\,\hat{v}_s/4)\,(\Omega_{\rm cr}/\Gamma_L)\,(e'_{\rm cr} + P'_{\rm cr})/e_B$) when linear damping dominates and $e_+ \approx e_- \sim (S_{et}\,e_B/\Gamma_{NL})^{1/2}$ when non-linear damping dominates (which occurs when $4\,\gamma_{NL}\,s_{et} \gtrsim (\gamma_L + \tilde{h})^2$).

In the SC limit ($|\Phi| \gg 1$): the streaming speed $\bar{v}_A/v_{A,\rm eff} = (x_+ - x_-)/(x_+ + x_-) \rightarrow -{\rm sign}(\tilde{g})$, so $\bar{v}_A$ is just the effective Alfvén speed ($\pm v_{A,\rm eff}$) directed down the CR pressure gradient. Only the $x_\pm$ aligned in this direction is large (the other vanishes), with the relevant $x \approx \bar{x} \sim (\gamma_L/2\,\gamma_{NL})\,(-1 + [1 + 4\,|\tilde{g}|\,\gamma_{NL}/\gamma_L^2]^{1/2})$ which corresponds to $e_A \sim S_{sc}^0/\Gamma_L$ (with $S_{sc}^0 \equiv |\mathbf{v}_A\cdot\nabla\cdot\mathbb{P}'_{\rm cr}|$) when linear damping dominates, and $e_A \sim (S_{sc}^0\,e_B/\Gamma_{NL})^{1/2}$ when non-linear damping dominates (which occurs when $4\,\gamma_{NL}\,|\tilde{g}| \gtrsim \gamma_L^2$).

B3 Justification of approximations

This allows us to formally justify some of the approximations used in the text to estimate scalings: if we write $S_\pm \sim S_{et} + S_{sc}^0$ as the 'total' driving and set $S_\pm$ equal to $Q_\pm \sim (\Gamma_L + \Gamma_{NL}\,e_A/e_B)\,e_A$ to solve for e_A (this was done in the text to justify our more approximate scalings), we obtain the correct qualitative behaviours of e_A/e_B in all the relevant limits above. The transition between ET and SC limits here, $S_{sc}^0 \gtrsim S_{et}$, corresponding to $|\tilde{g}|/s_{et}$, which usually determines Φ (though there can, in greater detail, be non-negligible corrections from the $\tilde{h}/\Gamma$ term in determining which limit is most relevant).

One interesting limit where this allows us to resolve some ambiguities is the case in highly neutral gas ($f_{\rm ion} \ll 1$), with gyroresonant Alfvén frequencies larger than the ion-neutral collision time, so in the expressions above $v_{A,\rm eff} \approx (|\mathbf{B}|^2/4\pi\,\rho_{\rm ion})^{1/2} = v_{A,\rm ideal}\,f_{\rm ion}^{-1/2} \gg v_{A,\rm ideal}$. In this case, S_{et} is suppressed by strong ion-neutral damping (which usually leads to $\Gamma_L \sim \Gamma_{\rm in} \gg \Gamma_{NL}$), while $\tilde{g}^2 \propto 1/f_{\rm ion}$ and $\tilde{h} \propto 1/f_{\rm ion}^{1/2}$ are enhanced, so for conditions of relevance in e.g. GMCs this means $|\Phi| \gg 1$ and the system rapidly converges to the SC regime with streaming at $v_{A,\rm eff} \propto 1/f_{\rm ion}^{1/2}$, essentially independent of the strength of $\nabla e'_{\rm cr}$ or S_{et} on larger scales.

We can also return to the approximations regarding time-scales made above. In equation (B1), we see from our steady-state solutions that the $c^2\,\nabla\cdot\mathbb{P}'_{\rm cr}$ term is never negligible (it acts as a source term), while the relative importance of the $F'_{e,\rm cr}$ and $H'_{\rm cr}$ terms depends on whether the flux is super or sub-Alfvénic. In any case, noting that the $F'_{e,\rm cr}$ term can be written $D_t\,F'_{e,\rm cr} \sim \bar{\nu}_s\,F'_{e,\rm cr} + ...$, we immediately confirm that the equation is driven towards steady-state on the very short scattering time-scale $\sim \bar{\nu}_s^{-1} \sim 30\,{\rm yr}\,R_{\rm GV}^{1/2}$ (for empirically favoured $\bar{\nu}_s$ values). In equation (B2), in steady-state, the dominant driving ($S_{et,\pm}$ or SC $F'_{e,\rm cr} - H'_{\rm cr}$ term) terms have magnitude of order the damping terms $\Gamma_\pm\,e_\pm$, so the equation $D_t e_\pm = -\Gamma_L\,e_\pm + ...$ is driven to steady-state on the local damping time $\sim \Gamma_\pm^{-1} \sim (30-300)\,{\rm yr}\,(R_{\rm GV}/B_{\mu\rm G})^{1/2}\,(10\,{\rm km\,s^{-1}}/v_{A,\rm eff})$ (using the scalings from Section 2.5 for $\Gamma_{\rm turb/LL}$ and $\Gamma_{\rm dust}$, assuming typical LISM properties; if other damping terms are also important, then $\Gamma_\pm^{-1}$ will be even smaller). This is similar to the scattering time. The 'gradient terms' $\mathcal{O}(\nabla[u_{\rm gas},\,v_{A,\rm eff}]\,e_\pm)$ are smaller than the other terms in equation (B2) by a factor $\mathcal{O}(|\nabla[u_{\rm gas},\,v_{A,\rm eff}]|/\Gamma_\pm) \sim 10^{-4}\,(R_{\rm GV}/B_{\mu\rm G})^{1/2}\,\ell_{\nabla,\rm ISM,10}^{-1}$ where $\ell_{\nabla,\rm ISM,10} = \ell_{\nabla,\rm ISM}/10\,{\rm pc}$ with $\ell_{\nabla,\rm ISM}$ the gradient scale-length of the bulk ISM properties ($\mathbf{u}_{\rm gas}$ or $v_{A,\rm eff}$), justifying their neglect above. These time-scales $\mathcal{O}(1/\nabla[u_{\rm gas},\,v_{A,\rm eff}])$ are of course also the same as the characteristic time-scales for bulk ISM properties to evolve (e.g. $v_{A,\rm eff}$, $\mathbf{B}$, $\mathbf{u}_{\rm gas}$, ρ from the usual MHD equations). Thus, this justifies our assumption that these MHD properties can be taken as approximately constant over the time-scale for equations (B1)–(B2) to reach local steady-state.

Now consider the CR energy equation $D_t e'_{\rm cr} = ...$, assuming the CR flux $F'_{e,\rm cr}$ and $e_\pm$ equations have reached local steady-state [equation (11) in the text]. The source/sink term $\mathcal{S}'_{\rm other,\,cr}$ is small compared to other terms at rigidities $\gtrsim$GV (except in special environments, e.g. at sources). Examination shows that the 'diffusive re-acceleration' term $\propto (v_{A,\rm eff}^2 - \bar{v}_A^2)/c^2$ is always small: it vanishes identically in the SC limit, but even in the ET limit it is suppressed by $\mathcal{O}(v_{A,\rm eff}^2/c^2)$ for any plausible $\bar{\nu}_s$ (see Hopkins et al. 2021a). The 'streaming' and 'adiabatic' terms ($\sim \nabla\cdot(\bar{v}_A\,e'_{\rm cr}\,\hat{\mathbf{b}})$ and $\sim P'_{\rm cr}\,\nabla\cdot(\mathbf{u}_{\rm gas} + \bar{v}_A\,\hat{\mathbf{b}})$) involve the same 'gradient' or ISM bulk-property time-scales $\mathcal{O}(1/\nabla[u_{\rm gas},\,v_{A,\rm eff}])$ as defined above. The term in the $D_t e'_{\rm cr}$ equation that can evolve most rapidly is the 'diffusive' term $\sim \nabla\cdot[\kappa_\parallel\,\hat{\mathbf{b}}\hat{\mathbf{b}}\cdot\nabla e'_{\rm cr}]$ (with $\kappa_\parallel \equiv v_{\rm cr}^2/3\,\bar{\nu}_s$), which drives $e'_{\rm cr}$ to equilibrium on the diffusive time-scale $\sim \ell_{\nabla,\rm cr}^2/\kappa_\parallel$ with $\ell_{\nabla,\rm cr} \sim e'_{\rm cr}/|\nabla e'_{\rm cr}|$ the CR energy gradient scale length. This is larger than the CR scattering time $\bar{\nu}_s^{-1}$ for $\ell_{\nabla,\rm cr} \gtrsim c/(3\,\bar{\nu}_s) \sim 3\,{\rm pc}\,R_{\rm GV}^{1/2}$ (i.e. $\ell_{\nabla,\rm cr}$ larger than the CR scattering mean free path). Since we observe and expect $\ell_{\nabla,\rm cr} \gtrsim$ kpc, we confirm that $e'_{\rm cr}$ will generically converge to steady-state on much longer time-scales than $F'_{e,\rm cr}$ or $e_\pm$.

[24] The solution here assumes $\Gamma_L > 0$, i.e. linear damping. If instead there were net linear driving from a driving source not considered here, so $\Gamma_L < 0$ in our language, then the physical solution branch for γ_{NL} small and $\gamma_L < -\tilde{h}$ becomes $\bar{x} \rightarrow s_{et}\,(1-\sqrt{1+\Phi^2})/(\gamma_L + \tilde{h})$.

[25] The ϕ term here accounts for the fact that diffusive re-acceleration produces a net transfer of energy from the scattering modes (e_A) to the CRs ($e'_{\rm cr}$) when $e_+ \approx e_-$ (in the ET limit), so acts like an additional linear damping term even when $\Gamma_L \rightarrow 0$.

APPENDIX C: EXTRINSIC TURBULENCE: BASIC SCALINGS AND RESULTS

It is helpful to briefly review some properties of ET. In a turbulent cascade with velocity and magnetic field fluctuations $\delta\mathbf{v}$, $\delta\mathbf{B}$ on a scale $\lambda \sim 1/k$, most of the energy ($|\delta\mathbf{v}^2(k)| \sim k\,\mathcal{E}(k)$) is concentrated around the driving scale $\lambda \sim \ell_{\rm drive}$ ($\gtrsim 0.1-1$ kpc in the ISM/CGM), with Alfvén Mach number $\mathcal{M}_A \sim \langle|\delta\mathbf{v}^2(\lambda\sim\ell_{\rm drive})|\rangle^{1/2}/v_{A,\,\rm ideal} \gtrsim 1$. On the largest super/trans-sonic/Alfvénic scales, this can give rise to a compressible and weakly pressurized Burgers (1973)-like power-spectrum ($|\delta\mathbf{v}^2(\lambda)| \propto \lambda$; Schmidt et al. 2009; Konstandin et al. 2012; Hopkins 2013; Squire & Hopkins 2017). Below the Alfvén scale ℓ_A, where $|\delta\mathbf{v}^2(\lambda\sim\ell_A)|\rangle^{1/2} \sim v_{A,\,\rm ideal}$ the fluctuations are sub-Alfvénic, by definition. In the Galactic ISM, typically $\ell_A \sim 10-100$ pc (Elmegreen 2002; Mac Low & Klessen 2004). The defining feature of a traditional strong inertial-range cascade is the energy condition, $S_{\rm turb} \sim e_{\rm turb}/\tau_{\rm cas} \propto |\delta\mathbf{v}(\lambda)|^2/\tau_{\rm cas}(\lambda) \sim$ constant, where $\tau_{\rm cas}$ is the cascade or energy-transfer or decoherence time, which can be parametrized over the inertial range as $\tau_{\rm cas} \sim (\ell_A/v_{A,\,\rm ideal})\,(\lambda/\ell_A)^{\alpha}$, with some $\alpha > 0$ (with $\alpha \leq 1$ almost always required).[26] For now, we neglect the difference between the parallel (to $\hat{\mathbf{b}}$) $k_\parallel$ and perpendicular $k_\perp$ components of k, but recall that what we need to calculate CR scattering rates is $e_{\rm turb}(k_\parallel)$, since it is the parallel component $k_\parallel$ that controls the scattering terms (e.g. Voelk 1975; and for gyroresonance, $k_\parallel \sim 1/r_{g,\,\rm cr}$). If there is strong damping/dissipation, then the dissipation/Kolmogorov scale of the turbulence $k_{\rm diss} \sim 1/\lambda_{\rm diss}$ occurs when some dissipation/damping rate $\sim\Gamma(k,\,\ldots)\,|\delta\mathbf{v}|^2$ becomes larger than the turbulent dissipation/cascade/transfer rate $\sim|\delta\mathbf{v}|^2/\tau_{\rm cas}$, i.e. $\Gamma \gtrsim 1/\tau_{\rm cas}$. For example, for some kinematic viscosity $\Gamma_{\rm visc} \sim \nu_{\rm visc}\,k^2$, we have $\lambda_{\rm diss} \sim \ell_A\,(\ell_A\,v_{A,\,\rm ideal}/\nu_{\rm visc})^{-1/(2-\alpha)}$ (i.e. $\ell_A\,v_{A,\,\rm ideal}/\nu_{\rm visc}$ is the Reynolds number).

The gyroresonant scale $\lambda_g \sim 1/k_g \sim r_{g,\,\rm cr} \sim 10^{-6}\,{\rm pc}\,R_{\rm GV}/B_{\mu\rm G}$, so $\lambda_g \ll \ell_A$ at energies of interest, hence $|\delta\mathbf{v}(\lambda_g)| \ll v_{A,\,\rm ideal}$ and $|\delta\mathbf{B}(\lambda_g)| \ll |\mathbf{B}|$, but we also know this empirically, since the observationally constrained CR scattering rates require $|\delta\mathbf{B}|/|\mathbf{B}| \sim 3\times10^{-4}\,R_{\rm GV}^{0.2}$. This means we can at least approximately decompose the fluctuations into a linear superposition of Alfvén, slow, and fast magnetosonic modes, with $|\delta\mathbf{B}|/|\mathbf{B}| \sim |\delta\mathbf{v}|/v_{A,\,\rm ideal} \ll 1$, and we can treat the gas as weakly compressible and smooth (gradient length scales of bulk ISM properties are much larger than λ_g).

C1 Anisotropy: Alfvénic and slow cascades

First consider the Alfvénic case. Alfvén waves are generally weakly damped by collisionless processes in ionized gas down to the ion gyro scale, so assume we can temporarily neglect damping. Since these are incompressible modes, we can rewrite the MHD equations in terms of the convenient Elsasser variables:

$$\partial_t\mathbf{Z}^+ - (\mathbf{v}_A\cdot\nabla)\mathbf{Z}^+ + (\mathbf{Z}^-\cdot\nabla)\,\mathbf{Z}^+ = -\nabla p/\rho$$
$$\partial_t\mathbf{Z}^- + (\mathbf{v}_A\cdot\nabla)\mathbf{Z}^- + (\mathbf{Z}^+\cdot\nabla)\,\mathbf{Z}^- = -\nabla p/\rho \qquad \text{(C1)}$$

where $\mathbf{v}_A = v_A\,\hat{\mathbf{b}}$, $\mathbf{Z}^+ \equiv \delta\mathbf{v} + \delta\mathbf{B}/(4\pi\rho)^{1/2}$, $\mathbf{Z}^- \equiv \delta\mathbf{v} - \delta\mathbf{B}/(4\pi\rho)^{1/2}$.

There are two possible limits to equation (C1). In limit **(1)**, the non-linear term ($(\mathbf{Z}^-\cdot\nabla)\,\mathbf{Z}^+$ or $(\mathbf{Z}^+\cdot\nabla)\,\mathbf{Z}^-$) is small. If this term is negligible, then we trivially recover the equations for Alfvén wave packets without any interactions: i.e. the equations do not feature any 'cascade' per se, but simply admit whatever spectrum of Alfvén waves we wish to externally impose, by introducing some other source term (e.g. SC driving or our proposed novel driving mechanisms). If the non-linear term is not completely ignored but still small (e.g. $\langle|(\mathbf{Z}^-\cdot\nabla)\,\mathbf{Z}^+|^2\rangle \ll \langle|(\mathbf{v}_A\cdot\nabla)\mathbf{Z}^+|^2\rangle$), then we obtain the classic assumptions of 'weak' Alfvénic turbulence. The conditions where this might occur in practice (in the absence of some other small-scale driving) are restrictive (see Lazarian 2016), but there is a bigger problem. As shown elegantly in Sridhar & Goldreich (1994; see also Schekochihin 2022 for a pedagogical presentation), an *isotropic* IK-type weak 'cascade' as envisioned by e.g. Kraichnan (1965) cannot exist (it is neither physically nor mathematically self-consistent): instead, the weak cascade occurs purely along $k_\perp$, conserving $k_\parallel$, so there is again *no* cascade to define $e_{\rm turb}(k_\parallel)$ nor is there any connection between $e_{\rm turb}(k_\parallel)$ for different $k_\parallel$ (weak Alfvénic turbulence simply redistributes this energy to different $k_\perp$ at the same $k_\parallel$, which has no effect to leading order on CR scattering). In other words, we once again simply recover whatever Alfvén spectrum $e_{\rm turb}(k_\parallel)$ we choose to impose by introducing some *other*, non-ET source term.

In limit **(2)**, the non-linear term is not negligible (e.g. $\langle|(\mathbf{Z}^-\cdot\nabla)\,\mathbf{Z}^+|^2\rangle \gtrsim \langle|(\mathbf{v}_A\cdot\nabla)\mathbf{Z}^+|^2\rangle$). In this limit, a cascade linking $\mathcal{E}$ at different $k_\parallel$ is possible, and making additional assumptions leads to, for example, the specific 'strong' turbulence cascade of Goldreich & Sridhar (1995; GS95, or variations proposed in Boldyrev 2005 or others reviewed in Schekochihin 2022), all of which give $\mathcal{E} \propto k_\parallel^{-1}$, i.e. $\delta_{\rm s} = 0$, as noted in the main text. More generally, a simple argument that $\delta_{\rm s}$ must be ≤ 0 in this regime goes as follows. Define the parallel scale of a mode as $l_\parallel \sim 1/k_\parallel$, such that $\mathcal{O}[(\mathbf{v}_A\cdot\nabla)\mathbf{Z}^+] \sim v_A\,Z_\lambda/l_\parallel$, and note that since $|\delta\mathbf{v}| \ll v_A$, this limit **(2)** requires $l_\parallel \gtrsim l_\perp\,(v_A/|\delta\mathbf{v}|) \gg l_\perp$ [or else the non-linear term would again be negligible, putting us in limit **(1)**]. This means $k \approx k_\perp$, so $\lambda \sim 1/k = l_\perp$. Without loss of generality, define $l_\parallel \propto v_A\,\tau_{\rm cas}\,(\ell_A/l_\parallel)^{\alpha_\parallel}$ over some dynamic range, such that $\mathcal{O}[(\mathbf{v}_A\cdot\nabla)\mathbf{Z}^+] \sim \mathcal{O}[(l_\parallel/\ell_A)^{\alpha_\parallel}\,Z_\lambda/\tau_{\rm cas}]$. Trivially, $\alpha_\parallel \geq 0$ is required so that the linear term is equal to or smaller than the non-linear & 'cascade' terms [otherwise, if $\alpha_\parallel < 0$, for $l_\parallel \sim \lambda_g \ll \ell_A$ we would immediately arrive back in limit **(1)**]. Note that the critical balance assumption corresponds specifically to $\alpha_\parallel = 0$. Now, we can also allow for some arbitrary losses from the cascade across scales by defining the cascade rate $S \sim e_{\rm turb}/\tau_{\rm cas} \sim S_0\,(l_\parallel/\ell_A)^{\alpha_S}$, where again any physical cascade requires $\alpha_S \geq 0$ (an un-damped cascade corresponds to $\alpha_S = 0$, but non-zero dissipation or losses can decrease the energy on smaller scales). Now if we recall $\nu \propto \Omega\,|\delta\mathbf{B}(k_\parallel\sim 1/r_{g,\rm cr})|^2/|\mathbf{B}|^2 \sim k_\parallel\,e_{\rm turb}(k_\parallel)|_{k_\parallel\sim 1/r_{g,\rm cr}} \sim S_0\,(l_\parallel/\ell_A)^{\alpha_S}\,\tau_{\rm cas}/l_\parallel \propto l_\parallel^{\alpha_\parallel+\alpha_S}$, we have $\delta_{\rm s} = -(\alpha_\parallel+\alpha_S) \leq 0$.

In short, it is not possible to construct an internally consistent Alfvénic *cascade* with $\delta_{\rm s} > 0$. Anisotropy in the form of critical balance with an un-damped Alfvénic cascade gives $\delta_{\rm s} = 0$. Adding losses/dissipation at scales between gyroresonant and driving only further decreases $\delta_{\rm s}$. Violating the critical balance-type assumptions (by e.g. introducing a non-zero $\alpha_\parallel$, in our notation above) leads to one of two outcomes. **(1)** There is no 'cascade' or any interaction between modes with different parallel wavenumbers (if $\alpha_\parallel < 0$), and the power $|\delta\mathbf{B}(k_\parallel)|^2$ must be set not by ET but by some other source term driving modes independently on each scale. Or **(2)** the cascade produces $\delta_{\rm s} < 0$ if $\alpha_\parallel > 0$, i.e. if the anisotropy is even larger than required for critical balance.[27]

[26] From the energy condition, this immediately gives $e_{\rm turb}(k) \propto k^{-\alpha}$, or the 1D $\mathcal{E}(k) \propto k^{-(1+\alpha)}$. So, e.g. the commonly cited 'K41-like' ($\mathcal{E}(k) \propto k^{-5/3}$), 'IK-like' ($\mathcal{E}(k) \propto k^{-3/2}$), and Burgers (1973)-like ($\mathcal{E}(k) \propto k^{-2}$) isotropic power spectrum scalings correspond to $\alpha = (2/3,\ 1/2,\ 1)$, respectively.

[27] Of course, as many have pointed out, any Alfvénic cascade THAT does not obey critical balance will be pushed (by a weak cascade or causality/de-correlation) towards a state of critical balance. We simply wish to stress that

For more rigorous discussion, we refer interested readers to Schekochihin (2022) for a review of more detailed Alfvénic turbulence models, demonstrating that even models that are imbalanced, intermittent, decaying, damped, or otherwise strongly modified, all obey $\delta_s \leq 0$ in our language.

Note that, as many others have pointed out, slow modes are subject to a similar anisotropy constraint to Alfvén waves as described above (which again leads to $\delta_s \leq 0$), *and* are subject to additional magnetosonic damping terms, which further constrain $\delta_s \leq 0$ as we discuss for fast modes below (see e.g. Cho & Lazarian 2003; Yan & Lazarian 2004; Schekochihin et al. 2009, and references therein). Thus, we do not discuss them further.

C2 Fast modes

Now consider instead a fast magnetosonic cascade. It is at least theoretically possible, in principle, that *within the inertial-range* these could produce an isotropic cascade ($k_\parallel \sim k_\perp \sim k$) with the desired scaling of $e_{\rm turb}(k) \propto k^{-\alpha}$: if e.g. $\delta_s \sim 0.6$ is observationally required, this would imply an inertial-range $\tau_{\rm cas} \sim \lambda^{0.4}$ or $|\delta \mathbf{v}| \propto \lambda^{0.2}$ ($\alpha \sim 0.4$). But it is important to stress that even the inertial-range behaviour on small scales ($\ll \ell_A$) is not theoretically clear: while e.g. Cho & Lazarian (2003) and Ferrand et al. (2020) argue for a spectrum with a Zakharov & Sagdeev (1970)-type weak cascade $e_{\rm turb}(k) \propto k^{-1/2}$ below the sonic/Alfvén scale (which would give $\delta_s \sim 0.5$ in the inertial range, within the observationally allowed range), others have argued from both analytic theoretical grounds (Kadomtsev & Petviashvili 1973; Elsässer & Schamel 1976; Shivamoggi 1992; Galtier et al. 2000; Kuznetsov & Krasnoselskikh 2008; Galtier & Banerjee 2011; Shivamoggi 2011; Sun 2016) and numerical simulations (Elsasser & Schamel 1974; Erlebacher et al. 1990; Mee & Brandenburg 2006; Kowal & Lazarian 2010; Lee et al. 2010; Makwana & Yan 2020) that the spectrum should be closer to $e_{\rm turb}(k) \propto k^{-1}$ (giving $\delta_s = 0$). And the classic Kolmogorov (1941) (K41)-type scaling $e_{\rm turb}(k) \propto k^{-2/3}$ ($\delta_s = 1/3$ in the inertial range), though often cited in older 'leaky box' models for CR transport that did not include a scattering halo, actually provides a poor fit to the observations in modern models that include any extended scattering halo (see e.g. Blasi & Amato 2012; Vladimirov et al. 2012; Gaggero et al. 2015; Cummings et al. 2016; Guo et al. 2016; Jóhannesson et al. 2016; Korsmeier & Cuoco 2016; Evoli et al. 2017; Amato & Blasi 2018; De La Torre Luque et al. 2021; Hopkins et al. 2021a)

It is also not clear that isotropy is a good assumption on small scales even for fast modes (see e.g. Kuznetsov & Krasnoselskikh 2008; Lee et al. 2010; Brandenburg & Nordlund 2011, and references therein). If there is significant anisotropy, for reasons similar to those above, it will generically tend to decrease δ_s.

But as discussed in the text, an entirely un-ambiguous problem is that isotropic fast modes at gyroresonant scales are very strongly damped. Even in a fully ionized medium, collisionless damping of fast modes is orders of magnitude more efficient than for Alfvén modes on these scales.[28] Depending on the assumptions of ISM/CGM properties, mode angles, and the cascade time-scale, if we define the 'damping scale' $k_{\rm diss} \sim 1/\lambda_{\rm diss}$ as that where for *some* mode angle $\cos\theta \equiv k_\parallel/k$, the most-rapid fast-mode damping rate $\Gamma_{\rm fast}$ is larger than the cascade rate $1/\tau_{\rm cas}$, we would obtain $\lambda_{\rm diss} \sim 10^{-4} - 10\,{\rm pc}$ (see e.g. fig. 1 and table 1 of Yan & Lazarian 2004). More importantly, accounting for the combination of viscous, collisionless, and neutral damping (with realistic ISM/CGM scalings), it is essentially impossible to make $\lambda_{\rm diss}$ smaller than $r_{g,\,\rm cr}$ at rigidities $< 100 - 1000\,{\rm GV}$.[29] As a result, we argued in the text that $\delta_s \leq 0$. Detailed numerical calculations showing $\delta_s \lesssim 0$ is always the case for all $R_{\rm cr} \lesssim 100 - 1000\,{\rm GV}$ for fast-mode ET, accounting in detail for exact expressions of the scattering rates and their detailed dependence on pitch angle, mode angle, and wavelength, along with the full range of angle-dependent damping rates from different processes (following more exact integral expressions for CR scattering physics), have been extensively presented, including in YL04 (their fig. 2), Yan & Lazarian (2002, 2004, 2008), and Kempski & Quataert (2021). And of course our fast-mode ET model in the main text (Fig. 4; *right*) is one such calculation as well. So here we only seek to justify this heuristically (see Kempski & Quataert 2021 as well for a similar discussion).

First, consider the effects of damping on gyroresonant CR scattering. If the damping is isotropic (as with e.g. neutral viscosity or ion-neutral damping per Spitzer 1978, in regions with neutral fraction $f_{\rm neutral} \gtrsim 0.001 - 0.01$; see text and Hopkins et al. 2021d), the spectrum is, by definition, truncated at $k_\parallel \gtrsim k_{\rm diss}$ exponentially or super-exponentially,[30] equivalent to $\delta_s \ll 0$. But even in a fully ionized medium with $\beta_{\rm plasma} \ll 1$ assuming the dominant damping is e.g. from Braginskii viscosity or collisionless damping, which are anisotropic and do not damp parallel modes, we have a damping rate of the form: $\Gamma_{\rm fast} \propto k^{1+\alpha_k}\,\sin^2(\theta)$, where $0 \leq \alpha_k \leq 1$ depends on e.g. whether collisionless or viscous damping dominates (and $\beta_{\rm plasma}$). At scales $\lambda < \lambda_{\rm diss}$, modes with $\theta > \theta_c$ will have $\Gamma_{\rm fast} \gg 1/\tau_{\rm cas}$ and be truncated, so if we make the most optimistic assumption that the

$$\nu_{\rm visc,\,eff} \equiv \nu_{\rm visc,\,ion,\,0}\, f_{\rm v,\,ion}(\theta)\, f_{\rm ion} + \nu_{\rm visc,\,neutral}\, f_{\rm neutral} \quad \text{(C4)}$$

$$\nu_{\rm visc,\,ion,\,0} \sim 0.6\times10^{18}\,{\rm cm^2\,s^{-1}}\,T_4^{5/2}\,n_1^{-1} \quad \text{(C5)}$$

$$\nu_{\rm visc,\,neutral} \sim 3\times10^{20}\,{\rm cm^2\,s^{-1}}\,T_4^{1/2}\,n_1^{-1} \quad \text{(C6)}$$

$$f_{\rm v,\,ion}(\theta) \approx \begin{cases} \sin^2(\theta) & (\beta_{\rm plasma} \ll 1) \\ |1-3\,\cos^2(\theta)|^2 & (\beta_{\rm plasma} \gg 1) \end{cases} \quad \text{(C7)}$$

Where for small θ, at $\beta_{\rm plasma}$ not too large $f_{\rm fast,\,L} \equiv (\omega_{\rm fast}/k\,v_{\rm fast})\,(\sqrt{\pi\,\beta_{\rm plasma}}/4)\,\sqrt{m_e/m_p}\,\exp[-m_e/m_p\,\beta_{\rm plasma}\,\cos^2(\theta)]$ (Ginzburg & Syrovatsky 1961), while for very large $\beta_{\rm plasma}$, $f_{\rm fast,\,L} = (2/\cos^2(\theta))\,(\omega_{\rm fast}/k\,v_{\rm fast})\,(\omega_{\rm fast}/\omega_{\rm c,\,i})$ (with $\omega_{\rm fast}$ the fast-mode frequency at wavenumber k, and $\omega_{c,\,i}$ the ion cyclotron frequency; Foote & Kulsrud 1979). Note that viscous damping in ionized gas with $\beta_{\rm plasma} \gtrsim 1$ acts similar to isotropic (neutral) damping, in that it strongly damps parallel fast waves. In evaluating the full CR scattering rate expressions, this has the same practical effect of strongly truncating the gyroresonant scattering term (giving $\delta_s \ll 1$).

even transient violations of this condition fail to produce an Alfvénic cascade with $\delta_s > 0$.

[28]The dominant fast-mode damping terms (in addition to the weaker terms in the main text which also apply to Alfvén waves) are: viscous damping $\Gamma_{\rm fast,\,visc}$ (both by neutrals and Braginskii 1965 viscosity from ions) and collisionless/Landau damping $\Gamma_{\rm fast,\,L}$:

$$\Gamma_{\rm fast,\,visc} \equiv k^2\,\nu_{\rm visc,\,eff} \quad \text{(C2)}$$

$$\Gamma_{\rm fast,\,L} \equiv \frac{\sin^2(\theta)}{\cos\theta}\,k\,v_{\rm fast}\,f_{\rm fast,\,L} \quad \text{(C3)}$$

[29]Given the most optimistic possible assumptions for reducing $\lambda_{\rm diss}$, it may be possible in some phases of the ISM, such as the WIM, to make $r_{g,\,\rm cr} > \lambda_{\rm diss}$ at $>100\,{\rm GV}$, while more typical assumptions for the WIM and even the most optimistic assumptions for the WNM and CNM or GMCs require $\gtrsim 1000\,{\rm GV}$. In hot gas in the Galactic coronae, HIM, and CGM/halo, $\lambda_{\rm diss}$ becomes much larger, and it is plausible that $\lambda_{\rm diss} \gtrsim r_{g,\,\rm cr}$ up to $>10^6$ GV (i.e. up to PeV CR energies).

[30]Even if we assumed the mathematically 'weakest possible' cutoff for the spectrum below the damping scale, i.e. we assume $S \propto S_0\,(k_{\rm diss}/k)^{\alpha_S}$ continues for $k > k_{\rm diss}$, we must have $\alpha_S > 0$. Equating this driving with the (by definition dominant) neutral damping $Q_{\rm visc,\,fast} = \Gamma_{\rm visc,\,fast}\,e_A \sim \nu_{\rm visc,\,n}\,k^2\,e_A$, we have $e_A \propto k^{-(2+\alpha_S)}$, i.e. $\delta_s = -1 - \alpha_s < -1$.

remaining modes simply continue their cascade uninterrupted, the surviving modes are confined to a narrow bicone with $|\theta| < \theta_c \ll 1$, where θ_c becomes smaller with increasing k. Equating $\Gamma_{\rm fast}$ and $\tau_{\rm cas}$ gives $\sin^2(\theta_c) \sim \theta_c^2 \sim \tau_{\rm cas}^{-1}(k)\, k^{-(1+\alpha_k)}$. So if $S \sim S_0 \sim$ constant on large scales is the total cascade power and begins (by assumption) isotropic, the power on smaller scales is necessarily reduced by a factor proportional to the solid angle of the undamped cone $\propto \theta_c^2$. Thus the energy of scattering modes with a given $k_\parallel$ must scale as: $e_A \propto S\, \tau_{\rm cas}\, \theta_c^2 \propto S_0\, k^{-(1+\alpha_k)}$, i.e. $\delta_{\rm s} = -\alpha_k \leq 0$. Note further that if there is any 'leakage,' i.e. transfer of energy between the weakly damped cone and broader mode angles that are rapidly damped, then S must decrease further even along the 'surviving' directions, so we take $S \rightarrow S_0\, (\lambda\, k_{\rm diss})^{\alpha_S}$ with $\alpha_S > 0$, giving $\delta_{\rm s} = -(\alpha_k + \alpha_{\rm s}) < 0$, and further reducing $\delta_{\rm s}$.

As pointed out in YL04 and others, if a spectrum is strongly suppressed or truncated at scales $\lambda_{\rm diss} \gg r_{g,\,\rm cr}$, then transit-time damping (TTD) from the larger-scale modes near $\lambda_{\rm diss}$ could still produce CR scattering, which dominates over the gyroresonant term. But for TTD, we must replace our gyroresonant expression from the main text ($\nu_{\rm s} \sim \Omega\, |\delta\mathbf{B}(k_\parallel)|^2/|\mathbf{B}|^2 \sim (v_{\rm cr}\, k_\parallel)\, (k_\parallel\, \mathcal{E}(k_\parallel))/e_{\rm B}$) with $\nu_{\rm s} \rightarrow \int (\Omega_{\rm cr}^2/|\mathbf{B}|^2) \mathcal{E}(k)\, dk\, |J'[k_\perp\, v_{{\rm cr},\,\perp}/\Omega_{\rm cr})]|^2\, 1/(k_\parallel\, v_\parallel) \mathcal{R}(k_\parallel\, v_\parallel - \omega\,|\,k,\, k_g,\, \ldots) \sim e_{\rm B}^{-1} \int_0^{k_{\rm diss}} dk \mathcal{E}(k)\, k\, v_{\rm cr}\, \mathcal{R} \sim (k_{\rm diss}\, v_{\rm cr})\, |\delta\mathbf{B}(k_{\rm diss})|^2/|\mathbf{B}|^2\, \mathcal{R}(k_g,\, k_{\rm diss},\, \ldots) \sim$ (constant) $\times\, \mathcal{R}$ from e.g. Voelk (1975). Here, $\mathcal{R}$ is some appropriate dimensionless 'response' or resonance function. Heuristically, this is just the statement that a CR is scattered in pitch angle by a random amplitude $|\Delta\mu| \sim |\delta\mathbf{B}(k)|/|\mathbf{B}|$ as it crosses a mode in time $\Delta t \sim \lambda/v_{\rm cr} \sim 1/k\, v_{\rm cr}$, so will random walk to an order-unity change in pitch angle after $N \sim |\mathbf{B}|^2/|\delta\mathbf{B}(k)|^2$ events, implying a scattering time $\nu_{\rm s}^{-1} \sim N^2\, \Delta t \sim [(k\, v_{\rm cr})\, |\delta\mathbf{B}(k)|^2/|\mathbf{B}|^2]^{-1}$. But if this is dominated by the integral over larger-scale modes, then it is by definition independent of $R_{\rm cr}$, so $\delta_{\rm s} = 0$. Moreover, if we account for any non-trivial response function $\mathcal{R}$ (describing how efficiently a mode of scale k can deflect the pitch-angle of a CR with gyroradius $r_{g,\,\rm cr} \sim 1/k_g$), it must be the case that $\mathcal{R}$ is a decreasing function of k_g/k for $k_g \gg k$, hence we must have $\delta_{\rm s} < 0$.

Finally, note that because of how the actual scaling of the spectrum S factors out in the above, and that these fast-mode damping mechanisms act on all scales of interest, the above conclusion that fast-mode damping requires $\delta_{\rm s} \leq 0$ applies not just to fast modes sourced by a larger-scale cascade, but *any* isotropically driven population of fast modes, even if they were driven or sourced around the gyroresonant scales.

C3 Summary

In summary, *it is not possible to construct a self-consistent 'cascade' model which produces $\delta_s > 0$* for CRs, as required (observed $\delta_{\rm s} \sim$ 0.5–0.7). We stress that the arguments above are quite generic: this is not a statement specific to one particular model of turbulence or to various controversial or uncertain assumptions. Rather, they arise from fundamental features of the MHD equations themselves (which require that any Alfvénic or slow 'cascade' linking different parallel $k_\parallel$ have $\delta_{\rm s} \leq 0$) or the fundamental nature of magnetosonic damping (which means any magnetosonic cascade where the most-rapidly damped-modes begin to be appreciably damped on a spatial scale $\lambda_{\rm diss}$ larger than the gyroradius $r_{g,\,\rm cr}$ at some $R_{\rm cr}$ must have $\delta_{\rm s} \leq 0$ at all smaller $R_{\rm cr}$).

There is, however, one rather straightforward way to provide the desired scattering: as we show in Section C1, if there exists some *other* source/driving term of Alfvén waves (other than a cascade related to any mode-coupling from larger or smaller scales) at parallel $k_\parallel$, and those waves are not extremely anisotropic [i.e. have typical $k_\parallel \gtrsim (|\delta\mathbf{v}(k_\parallel)|/v_A)\, k_\perp \gg 0.0003\, k_\perp$, so limit **(1)** in Section C1 applies], then it is perfectly allowed to construct an arbitrary spectrum $\mathcal{E}(k_\parallel) \propto k_\parallel^{\delta_{\rm s}-2}$ with the desired $\delta_{\rm s}$. These can be weakly damped (by e.g. the mechanisms in the text), and undergo a weak cascade mixing the perpendicular wavenumbers $k_\perp$ but leaving $\mathcal{E}(k_\parallel)$ unmodified, and satisfy all consistency constraints we discuss above. As noted in the text, standard SC theory would be one example of precisely this case, except that (for entirely different reasons) the form of the driving term $S_{\rm sc}$ produces the incorrect spectrum $\mathcal{E}(k_\parallel)$ (unless one also modifies the damping terms as we discuss). The other driving terms we propose: $S_{\rm new,\,lin}$ and $S_{\rm new,\,ext}$ also function in this manner.

At large CR rigidities, $\gtrsim 0.1 - 1$ TV, it becomes possible (at least in some ISM phases) to have $r_{g,\,\rm cr} \gtrsim \lambda_{\rm diss}$, so a 'traditional' ET-type theory can apply and at least in principle one could obtain a reasonable CR scattering rate from turbulence with an isotropic inertial-range spectrum with $k_\parallel\, \mathcal{E}(k_\parallel) \propto k_\parallel^{-1/2}$ (see e.g. Fornieri et al. 2021). But of course, it is also possible that additional source terms like those we propose could still be important on these scales as well.

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