Physics of Pair Producing Gaps in Black Hole Magnetospheres. II. General Relativity

Alexander Y. Chen1 and Yajie Yuan1,2

1 Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA; alexc@astro.princeton.edu
2 Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA; yyuan@flatironinstitute.org

Received 2019 August 19; revised 2020 April 20; accepted 2020 April 21; published 2020 June 3

Abstract

This is the second paper in a series where we examine the physics of pair producing gaps in low-luminosity accreting supermassive black hole systems. In this paper, we carry out time-dependent self-consistent fully general relativistic 1D particle-in-cell simulations of the gap, including full inverse Compton scattering and photon tracking. Similar to the previous paper, we find a highly time-dependent solution where a macroscopic vacuum gap can open quasiperiodically, producing bursts of $e^\pm$ pairs and high energy radiation. We present the light curve, particle spectrum, and photon spectrum from this process. Using an empirical scaling relation, we rescale the parameters to the inferred values at the base of the jet in M87, and find that the observed TeV flares could potentially be explained by this model under certain parameter assumptions.

Unified Astronomy Thesaurus concepts: Black hole physics (159); High energy astrophysics (739); Plasma astrophysics (1261); Nonthermal radiation sources (1119)

1. Introduction

It is widely believed that powerful jets from active galactic nuclei (AGN) are launched by rapidly spinning black holes through the Blandford–Znajek process (Blandford & Znajek 1977). When the black hole is threaded by a strong, coherent magnetic field, the rotating spacetime near the event horizon drags the field lines into rotation, resulting in a significant amount of Poynting flux flowing out along the magnetic field. The energy flow comes from the rotation of the black hole, and eventually gets dissipated further downstream along the jet, giving rise to the multiwavelength emission we observe from jets. The launching and acceleration of Poynting flux dominated jets have been demonstrated in general relativistic magnetohydrodynamic simulations (e.g., McKinney et al. 2012). However, it turns out that in the jet funnel, plasma tends to be evacuated either inward to the black hole or outward to infinity, and some source of continuous plasma supply is needed to carry the underlying electric current, which cannot be self-consistently captured in the MHD framework.

In reality, when the plasma supply runs low in the jet funnel, electrostatic gaps—regions with unscreened electric field—could develop and accelerate particles to high enough energies to initiate an $e^\pm$ pair cascade, which would replenish the plasma needed in the jet circuit, similar to what happens around pulsars (e.g., Blandford & Znajek 1977; Beskin et al. 1992; Hirotani & Okamoto 1998). But unlike pulsars where the pair producing photons are mainly generated by curvature radiation, near the black hole the dominant process is inverse Compton (IC) scattering of soft photons emitted by the accretion disk. The electrostatic gaps are probably most important in low-luminosity AGN. In high luminosity AGN, the accretion disk may already produce many MeV photons that could collide with each other and convert to pairs in the jet funnel. If these pairs are already enough to conduct the current, then a gap does not need to open up (e.g., Levinson & Rieger 2011).

The magnetospheric gap has been considered as a possible candidate site for producing the rapid $\gamma$-ray variability seen from certain AGN, especially radio galaxies (e.g., Levinson & Rieger 2011; Aleksic et al. 2014; Aharonian et al. 2017; Katsoulakos & Rieger 2018). The most promising source seems to be M87 (Katsoulakos & Rieger 2018). It has a massive black hole with mass $M \approx 6.5 \times 10^9 M_\odot$ (Event Horizon Telescope Collaboration et al. 2019), and an estimated jet power $L_{\text{jet}} \sim 10^{44} \text{erg s}^{-1}$ (e.g., Broderick & Tchekhovskoy 2015). Its very high energy ($\gamma$-ray) emission shows strong variability on timescales as short as a day (Abramowski et al. 2012). During the $\gamma$-ray flare, the VHE luminosity can reach $10^{42} \text{erg s}^{-1}$, and the flares have been observed to be accompanied by increased radio emission from the nucleus (Acciari et al. 2009), strongly suggesting particle acceleration in the vicinity of the black hole. In order to see whether the gap scenario can reproduce all the features of the $\gamma$-ray flares, a detailed study of the gap physics is needed.

So far in the literature there are quite a few works on the gap physics; most of them employ steady state/electrostatic models (Beskin et al. 1992; Hirotani & Okamoto 1998; Broderick & Tchekhovskoy 2015; Hirotani & Pu 2016; Hirotani et al. 2016, 2017; Pitsyna & Neronov 2016; Levinson & Segev 2017; Ford et al. 2018). However, the gap development and screening may be an intrinsically time-dependent process (Broderick & Tchekhovskoy 2015; Levinson & Segev 2017), similar to the polar cap gap in pulsars (Timokhin & Arons 2013). Recently, Levinson & Cerutti (2018) used 1D general relativistic particle-in-cell (GRPIC) simulations to study the discharge of a gap that is depleted of particles at the beginning. They found that except for the initial transient, the long term evolution reaches a quasi-steady state where no macroscopic gap is seen and pairs are generated everywhere randomly. Parfrey et al. (2019) carried out 2D GRPIC simulations of the jet launching process by immersing a rapidly spinning Kerr black hole in a uniform magnetic field. They used artificially simplified pair injection schemes, and found that the resulting plasma distribution depends sensitively on the pair creation physics.

In our previous paper, we also carried out 1D local simulations of the gap using a flat spacetime approximation (Chen et al. 2018, hereafter CY18). We found the gap development and screening to be a quasi-periodic process, a result qualitatively very different compared to Levinson & Cerutti (2018). We focused on the high
opacity scenario where inverse Compton scattering mainly happens in the Thomson regime, and constructed a simple analytic model to extrapolate numerical simulations to M87 and Sgr A*. It was found that for M87, the gap power is not enough to power the observed TeV flares. In this paper, we properly implement general relativity and a more complete numerical treatment for inverse Compton scattering. We extend our study into the deep Klein–Nishina (KN) regime and attempt to find an empirical scaling relation to extrapolate our simulation results. We will first describe our method comprehensively in Section 2, then present the main results in Section 3. We will comment on the implication of these results for M87 in particular in Section 4, and finish with a discussion of potential caveats and future extensions in Section 5.

2. Simulation Setup

2.1. Equations of Motion

We consider a nearly force-free magnetosphere around the black hole, with open field lines threading the event horizon, in a more or less axisymmetric, quasi-steady state. Somewhere along the field line, gaps may open up when particle density runs low. If the spatial and temporal scale of the gap is much smaller than the global characteristic scale (~a few \( r_g \equiv GM/c^2 \), the gravitational radius), the gap dynamics can be well described using 1D approximation along the field line.

In our numerical simulations, we take the deviation from the background force-free configuration to be our dynamic variable, and solve the equations along a 1D flux tube labeled by the flux function \( \psi(r, \theta) \) (see the Appendix for a description of the force-free configuration). The basic equations have a relatively simple form in the 3 + 1 formalism (e.g., Komissarov 2004). Here, the Kerr metric can be written as (we use Greek indices for 4-vectors, and Latin indices for spatial 3-vectors)

\[
dx^2 = (\beta^2 - \alpha^2)dr^2 + 2\beta\,dtdx^i + \gamma_{ij}dx^i dx^j, \tag{1}
\]

where \( \alpha \) is the lapse function and \( \beta \) is the shift vector. It is convenient to use Boyer–Lindquist coordinates as \( \gamma_{ij} \) is diagonal:

\[
\gamma_{ij} = \text{diag}(g_{rr}, g_{\theta\theta}, g_{\phi\phi}) = \text{diag}(\Sigma/\Delta, \Sigma, A \sin^2 \theta/\Sigma),
\]

where \( \Sigma = r^2 + a^2 \cos^2 \theta, \Delta = r^2 - 2Mr + a^2, \) and \( A = (r^2 + a^2)^2 - \Delta a^2 \sin^2 \theta \). Meanwhile,

\[
\alpha = \sqrt{\frac{g_{\theta\theta} - g_{\phi\phi}}{g_{\phi\phi}}} = \sqrt{\frac{\Delta \Sigma}{A}}, \tag{2}
\]

and \( \beta^i \) only has \( \phi \) component:

\[
\beta^\phi = -\omega, \tag{3}
\]

where \( \omega = -g_{\theta\phi}/g_{\phi\phi} = 2Mar/A \) is the angular velocity of the zero angular momentum observer (ZAMO). To alleviate the divergence near the event horizon, we use the tortoise coordinates which is a transformation from the Boyer–Lindquist coordinates by letting \( dx^i = dr/\Delta \), such that

\[
\xi(r) = \frac{1}{r_+ - r_-} \ln \left( \frac{r - r_+}{r - r_-} \right), \tag{4}
\]

\[
r(\xi) = \frac{r_+ - r_- e^{\xi(r_+ - r_-)}}{1 - e^{\xi(r_+ - r_-)}}, \tag{5}
\]

where \( r_\pm = M \pm \sqrt{M^2 - a^2} \) are the radius of the event horizons given by \( \Delta = 0 \). We can see that \( \xi \to -\infty \) as \( r \to r_+ \). The field equations that we solve in \((\xi, t)\) are:

\[
\frac{\partial}{\partial \xi}(K_i\partial \xi) = K_i(p - p_H), \tag{6}
\]

\[
\frac{\partial}{\partial t}(K_i\partial t) = K_i(j - j_H), \tag{7}
\]

\[
\frac{\partial}{\partial t}(K_i\rho) + \frac{\partial}{\partial \xi}(K_i j^\xi) = 0, \tag{8}
\]

where the factor \( K_i = \Delta \sqrt{\gamma}/(\partial_\theta \psi) \) accounts for field line divergence along the flux tube. We use the third equation to solve for the current due to particle movement; therefore, Gauss’s law (Equation (6)) is automatically satisfied as long as it holds initially. We always set \( \rho = \rho_H \) initially and \( D^\xi = 0 \), which trivially satisfies the Gauss’s law. We observe excellent charge conservation as even after millions of time steps charge density is always driven back to \( \rho_H \).

For particles, we assume they are well magnetized and can only move freely along magnetic field lines. In other words, their motion is constrained to 1D, and \( u^\xi = u'/\Delta \) completely determines the full 4-velocity:

\[
u^0 = \Omega u^\phi + \frac{B^\phi}{B^\xi} u^\xi, \tag{9}
\]

\[
u^\phi = \frac{B^\phi}{B^\xi} u^\xi, \tag{10}
\]

where \( B^\xi = B'/\Delta \), and \( \Omega \) is the angular velocity of the field line. We can then write down the Lagrangian in terms of coordinate \( \xi \) and velocity \( v_\xi = d\xi/dt \):

\[
L(t, \xi, v^\xi) = -\frac{1}{u^0} - \frac{q}{m} \Phi, \tag{11}
\]

where \( u^0 \) is written in terms of \( v^\xi \) as

\[
u^0 = \frac{1}{\sqrt{\alpha^2 - (S_2(v^\xi)^2 + 2S_1 v^\xi + S_3)}}, \tag{12}
\]

here

\[
S_1 = (\beta_\phi + \Omega_\phi)B^\phi/B^\xi, \tag{13}
\]

\[
S_2 = \gamma_\xi + \gamma_\phi B^\phi/B^\xi + \gamma_\theta (B^\theta/B^\xi)^2, \tag{14}
\]

\[
S_3 = \beta^2 + 2\Omega_\phi \beta_\phi + \gamma_\phi \Omega_\phi \Omega. \tag{15}
\]

The canonical momentum is

\[
p_\xi = \frac{\partial L}{\partial \dot{v}^\xi} = u^\phi (S_1 + S_2 v^\xi), \tag{16}
\]

and the Euler–Lagrange equations are

\[
\frac{dp_\xi}{dt} = \frac{\partial L}{\partial \dot{v}^\xi} = \frac{qE_\xi}{m} - \frac{u^0}{2} \frac{\partial \alpha^2}{\partial v^\xi} \tag{17}
\]

\[
+ \frac{u^0}{2} \left( \frac{\partial S_1}{\partial v^\xi} + 2\frac{\partial S_1}{\partial v^\xi} v^\xi + \frac{\partial S_2}{\partial v^\xi} (v^\xi)^2 \right), \tag{18}
\]
where \( E_\xi = -\partial_\xi \Phi \). The equations can also be written purely in terms of \( p_\xi \), using

\[
v_\xi = \frac{p_\xi / u^0 - S_1}{S_2},
\]

\[
u^0 = \sqrt{\frac{p_\xi^2 + S_2}{S_2(\alpha^2 - S_3) + S_2^2}}.
\]

In our field update equations \( D_\xi \) is the dynamic variable, while Equation (14) needs \( E_\xi \). We can obtain the latter through

\[
E_\xi = \alpha D_\xi = \alpha \gamma_\xi D_\xi.
\]

This set of equations naturally reproduces the inner and outer light surfaces, outside which particles flow either to the horizon or to infinity. We place the boundaries of our simulation box to be slightly outside the two light surfaces, so that all particles near the boundary can only outflow.

Photons are not confined to field lines so they obey the general equations of motion. In our example, the background field lines on the poloidal plane is more or less radial. It is a good approximation to neglect \( \theta \) component of the photon 4-velocity. Therefore the equations are

\[
\frac{du_\xi}{dt} = -\frac{u^0}{2} \partial_\xi \alpha^2 + u_\phi \partial_\beta \beta^0 - \frac{u_\xi^2 \partial_\xi \gamma_\xi^{\epsilon\xi} + u_\phi^2 \partial_\xi \gamma_\xi^{\phi\phi}}{2u^0},
\]

\[
\frac{du_\phi}{dt} = 0,
\]

\[
\frac{d\xi}{dt} = \gamma_\xi^{\epsilon\xi} \frac{u_\xi}{u^0},
\]

\[
u^0 = \frac{1}{\alpha} \sqrt{\gamma_\xi^{\epsilon\xi} (u_\xi)^2 + \gamma_\xi^{\phi\phi} (u_\phi)^2},
\]

where \( \gamma_\xi^{\epsilon\xi} = \gamma^{\epsilon\epsilon} / \Delta^2 \).

2.2. Radiative Transfer

We assume that the background soft photon field is isotropic in the ZAMO frame, and model the scattering event (IC and \( \gamma\gamma \) pair production) in this frame. For simplicity, we assume that the number density of the soft photons in the ZAMO frame is a power law independent of the radius:

\[
n(\epsilon) = \frac{n_0}{\epsilon_{\text{min}}} \left( \frac{\epsilon}{\epsilon_{\text{min}}} \right)^{-\alpha - 1}.
\]

We have chosen the normalization such that \( n_0 \) has the dimension of the total number of photons per volume. In addition, we consider the regime where the number density of the background soft photon field is much larger than those emitted by the particles in the gap, so that only scatterings on the background photons are important while nonlinear effects can be neglected.

The rate of scattering between an electron and the background photons in the ZAMO frame is (e.g., Blumenthal & Gould 1970)

\[
\nu_{\text{IC}} = \int \frac{1}{2} n(\epsilon) \sigma_{\text{IC,}\epsilon} c (1 - \beta \cos \theta) \sin \theta d\theta d\epsilon.
\]

Here \( \sigma_{\text{IC,}\epsilon} \) is the full KN cross section (e.g., Rybicki & Lightman 1986):

\[
\sigma_{\text{IC,}\epsilon} = \frac{3}{4} \frac{\sigma_T^2}{x^3} \left( \frac{2(x + 1)}{2x + 1} - \ln(2x + 1) \right) + \frac{\ln(2x + 1)}{2x} - \frac{3x + 1}{(2x + 1)^2},
\]

where \( x \) is the incident photon energy \( \epsilon' \) in the electron rest frame measured in units of the electron rest mass:

\[
\epsilon = \frac{\epsilon'}{m_e c^2} = \gamma (1 - \beta \cos \theta),
\]

and \( \gamma = \alpha u^0 \) is the energy of the electron in the ZAMO frame. At each time step, we determine the scattering probability \( P_{\text{IC}} \) using the following relation:

\[
P_{\text{IC}} = \nu_{\text{IC}} \Delta t = \nu_{\text{IC}} \alpha \Delta t,
\]

where \( \Delta t \) is the global time step, and \( \Delta \hat{t} \) its corresponding value in the ZAMO frame. If a scattering event occurs, we draw the scattered photon energy based on the IC spectrum (e.g., Jones 1968)

\[
\frac{dN}{dE_{\text{IC}}} = \frac{n(\epsilon)}{\epsilon} \frac{2\pi r^2 m_e c^3}{\epsilon} \times [2 \ln q + (1 + 2q)(1 - q) + \frac{1}{2} (\Gamma_{\epsilon} q)^2 (1 - q)],
\]

where \( E_1 \) is the energy of the scattered photon in units of the initial electron energy: \( \epsilon_1 = \gamma m_e c^2 E_1; \ Gamma = 4\epsilon c/(m_e c^2) \), \( q = E_1/\Gamma_{\epsilon} (1 - E_1) \). The range of \( E_1 \) is

\[
1 \gg \frac{\epsilon}{\gamma m_e c^2} \leq E_1 \leq \frac{\Gamma_{\epsilon}}{\Gamma_{\epsilon} + 1}.
\]

Since the photon is not tied to the field line, we compute its \( u_\xi \) and \( u_\phi \), when it is created, and follow its trajectory using the photon equations of motion in Section 2.1.

For \( \gamma\gamma \) pair production, the total cross section for photons of energy \( \epsilon, \epsilon \) and incident angle \( \theta \) is (Gould & Schréder 1967)

\[
\sigma_{\gamma\gamma} = \frac{1}{2} \pi r_\epsilon^2 (1 - \beta^2) \left[ (3 - \beta^4) \ln \frac{1 + \beta}{1 - \beta} - 2 \beta (2 - \beta^2) \right].
\]

where \( \beta c \) is the electron (positron) velocity in the center-of-mass frame,

\[
\beta = \left( 1 - \frac{1}{s} \right)^{1/2}, \quad \text{with } s = E_c (1 - \cos \theta). \tag{30}
\]

The threshold requires \( s > 1 \). A high energy photon of energy \( E \) traversing an isotropic background soft photon field with a number spectrum \( n(\epsilon) \) in the ZAMO frame will have an absorption probability per unit time

\[
\nu_{\gamma\gamma} = \int \frac{1}{2} n(\epsilon) \sigma_{\gamma\gamma} c (1 - \cos \theta) \sin \theta d\theta d\epsilon. \tag{31}
\]

When a photon converts to a pair, we set the direction of motion of the particles to be along the local \( B \) field, and the
photon energy is evenly split to the pair in the local ZAMO frame.

We tabulate all the cross sections and spectra and use a binary search table look-up to find the correct rates. This is an efficient procedure and we are able to process more than $10^9$ scatterings per second on a single GPU.

2.3. Units and Scales

We adopt a numerical unit system where $r_g$ is the unit of length, and $r_g/c$ is the unit of time. Magnetic field in numerical units is measured in terms of cyclotron frequency. We use a tilde to denote dimensionless quantities. For example, $\tilde{B} = eB r_g/(m_c c^2)$.

There are three main parameters of the problem. $B_0$ is the background magnetic field strength, which determines the background current $j_H$ and charge density $\rho_H$. It also sets a characteristic plasma skin depth

$$\lambda_p = \frac{c}{\omega_p} = \sqrt{\frac{m_e c^3}{4\pi\epsilon_0 e}} \sim \sqrt{\frac{m_e c^2 r_g}{B_0 e}},$$

which leads to $\tilde{B}_0 = (r_g/\lambda_p)^2$. In other words, $\tilde{B}_0$ sets the ratio between macroscopic and microscopic length scales.

The peak of the soft photon spectrum $\epsilon_{\text{min}}$ is the second parameter, which sets the energy scale of the radiative transfer. Particles with $\gamma \lesssim 0.1/\epsilon_{\text{min}}$ interact with the soft photon field in the Thomson regime, while those with $\gamma > 0.1/\epsilon_{\text{min}}$ fall into the KN regime and suffer from a reduced scattering cross section. Gamma-ray photons with energy $0.1/\epsilon_{\text{min}} \lesssim \gamma \lesssim 1/\epsilon_{\text{min}}$ have the smallest free path to $\gamma\gamma$ collision.

The third parameter $\tau_0$ is the characteristic optical depth to inverse Compton scattering in the Thomson regime. It scales with the number density $n_0$ of the soft photon field:

$$\tau_0 = r_g n_0 \sigma_T = \frac{r_g}{\ell_{\text{IC}}} = (\tilde{B}_0 \ell_{\text{IC}})^{-1},$$

where $\ell_{\text{IC}} = 1/(n_0 \sigma_T)$. $\tau_0$ determines the efficiency of IC scattering as well as $\gamma\gamma$ collision. If any gap develops in the magnetosphere, it is difficult to screen it before it grows to $h \sim r_g/\tau_0$, simply because that is the minimum mean free path the high energy photons need to travel before creating pairs. If $\tau_0$ is close to unity, then we expect the gap to at least grow to a scale of $r_g$.

We use the code Aperture for simulations in this paper. This version of the code is highly optimized on GPUs, and provides a unified framework for simulating the magnetospheres of neutron stars and black holes. The resolution for the runs in this paper range from 16,384 to 196,608 grid cells, with up to 200 million particles and photons, evolved over several million time steps. All of these runs were carried out on a single GPU, either a Tesla P100 or a Quadro P6000.

3. Results

3.1. Time-dependent Gap

In almost all cases, our simulations show a quasi-periodic gap opening and screening, very similar to the results of CY18. Typically the system will undergo an initial transient phase where a large gap opens, filling the box with plasma, then settle down to the quasi-periodic state. One such cycle is shown in Figure 1. Here a good criterion for forming a gap is based on the multiplicity, defined as the ratio between the actual number of charge carriers in the gap and the minimum amount needed to provide the current:

$$M = \frac{|\rho_+ - \rho_-| |e|}{\sqrt{\sigma_{\gamma\gamma} J}},$$

(34)

Whenever the local pair multiplicity drops below unity, parallel electric field starts to grow, accelerating particles and producing photons. It is remarkable that, except for the initial transient that depends on initial condition, the gap always grows from the null surface, where $\rho_H = 0$.

The quasiperiodic behavior of the gap can be seen from the integrated dissipation rate in the flux tube, $L = \int \alpha g_\gamma D' f' \sqrt{\gamma} \partial_t \psi \, dr$. The time evolution of this gap power can be interpreted as the light curve, and is plotted in Figure 2, where the power is normalized to the Poynting flux density in the same flux tube measured in the force-free simulation, $L_0 \approx 0.0124B_0^2$, which we can also identify as the jet power. Three runs of different $B_0$ and $\epsilon_{\text{min}}$ but the same $B_0 \epsilon_{\text{min}}$ and $\tau_0$ are plotted, and they have near identical time evolution and gap dissipation. As will be discussed in Section 3.2, we find that in general gap power and screening timescale only scale with the product $B_0 \epsilon_{\text{min}}$, but insensitive to the values of $B_0$ and $\epsilon_{\text{min}}$ themselves.

In addition to highly time-dependent gap power, we observe the same variability reflected in the photon flux near the outer light surface. We keep track of about 5% of all the particles and photons in the simulation, and look at all the photons within $0.5r_g$ of the outer light surface. The total energy as well as the spectrum of these outgoing photons vary according to gap activity. Figure 3 compares the spectra of the high state (when photon flux is highest) and low state (when photon flux is lowest). Both spectra form a power law, with the high state having a harder spectral index. The photon spectrum tends to extend a bit above 0.1/\epsilon_{\text{min}}, at which point it is strongly absorbed. The lower end of the spectrum is somewhat artificial, since we need to remove some low energy photons in the simulation in order to make it computationally feasible. In this particular run, all photons with energy below $10^3m_c c^2$ are removed at creation. However, photons created above this energy threshold are allowed to propagate and can be redshifted to lower energies if they travel from near the gap to the outer boundary. Even with lower energy photons removed, the energy escaping from the box is still dominated by photons, suggesting most of the gap power goes into radiation.

In our simulations the multiplicity of pairs escaping the box is usually between 10 and 50. However, most of the photons emitted in the gap do not convert in the box, and will continue to produce more particles if they travel outwards for a much larger distance, undergoing a post gap cascade, which has been discussed by, e.g., Broderick & Tchekhovskoy (2015). We expect the terminal multiplicity to be much higher than what we get in the local simulation.

3.2. Dependence on Parameters

In the high $\tau_0$ or small $\ell_{\text{IC}}$ case, e.g., $\tau_0 > 100$, IC scatterings mainly happen in the Thomson regime. The behavior of the gap is identical to what was reported in CY18. Despite the high
regime. Using the unit convention of \( \mathcal{L} \), the gap size is also almost \( \sim 100 \) mi, \( \sim 3 \) g, \( \sim \). Orange dashed line marks still holds in the high \( \gamma \) regime. Since the photon mean free path for \( \tilde{\gamma} \), not individually. With \( \tilde{\gamma} \) as an order 1 constants equation should only be valid in the Thomson regime, we neglect the geometric factors this run are \( \gamma_0 = 10, \tilde{B}_0 = 10^6 \), and \( \tilde{\gamma}_{\text{min}} = 10^{-3} \). The simulation box has 196,608 grid points. optical depth, the gap is macroscopic with a size \( h \sim 0.1-1r_g \), sometimes even larger during the initial transient. Particles are predominantly radiation reaction limited, and gap screening happens when energies of the upscattered photons are increased such that \( \tilde{\gamma}_r \sim h \). The fiducial optical depth, together with \( \tilde{\gamma}_{\text{min}} \), sets the energy scale at which gap screening occurs. The scaling relation derived in CY18, namely their Equation (16), still holds in the high \( \gamma_0 \) regime. Using the unit convention of this paper, the same relation can be written as (neglecting the order 1 constants)

\[
\tilde{B}_{\text{max}} \sim (\tilde{B}_0 \tilde{\gamma}_{\text{min}})^{-1/2} \tilde{\gamma}_{\text{min}}^{1/2 - \gamma_0}. \tag{35}
\]

Notice that this ratio depends only on the product of \( \tilde{B}_0 \) and \( \tilde{\gamma}_{\text{min}} \), not individually. With fixed \( \gamma_0 \) the gap size is also almost constant with respect to the product \( \tilde{B}_0 \tilde{\gamma}_{\text{min}} \). Although this equation should only be valid in the Thomson regime, we find that the dependence on the product \( \tilde{B}_0 \tilde{\gamma}_{\text{min}} \) is more universal. This is because particle energy only has one characteristic scale, which is \( \gamma_0 = m_e c^2 / \tilde{\gamma}_{\text{min}} \). Writing \( \tilde{\gamma} = \tilde{\gamma}_0 \), electric field \( E = \tilde{B}_0 \), particle number density \( n = \tilde{n} m_e c^2 / (e^2 \tilde{r}_g^2) \), the main equations become (neglecting the geometric factors) \( \partial_t \tilde{E} \sim \tilde{n}_+ - \tilde{n}_- \), and \( \partial_t \tilde{\gamma} \sim (\tilde{B}_0 \tilde{\gamma}_{\text{min}}) \tilde{E} \), which only depend on the product \( \tilde{B}_0 \tilde{\gamma}_{\text{min}} \).

When \( \gamma_0 \lesssim 100 \), particles are accelerated into the KN regime. Since the photon mean free path for \( \gamma-\gamma \) collision is smallest when the photon energy is comparable to \( m_e c^2 / \tilde{\gamma}_{\text{min}} \), the gap will be at least as large as the minimal \( \ell_{\gamma\gamma} \), which is \( \ell_{\gamma\gamma} \sim 5 \tilde{\gamma}_{\text{IC}} = 5r_g / \tilde{\gamma}_0 \). We observe that the gap size scales very weakly with parameters, only becoming larger when \( \tilde{B}_0 \tilde{\gamma}_{\text{min}} \) is so small that the gap voltage \( e \hbar \tilde{B}_{\text{max}} \) is comparable to \( m_e c^2 / \tilde{\gamma}_{\text{min}} \). Surprisingly, when \( \gamma_0 \lesssim 3 \), the particles are accelerated too far into the KN regime that the gap is no longer screened in our simulations.

The gap screening mechanism in the KN regime is different from what was described in CY18 for the Thomson regime. In the latter case IC loss is efficient so that most of the highest energy particles in the gap are radiation reaction limited, and the highest energy photons have the smallest free path. It is the highest energy photons that eventually produce the gap-screening pairs; therefore, we focused on the kinematics of the highest energy particles to derive the gap scaling relation. However, in the KN regime this is no longer the case. Cooling becomes inefficient when the particle Lorentz factor reaches \( 1 / \tilde{\gamma}_{\text{min}} \). Particles are accelerated to much higher energies than the radiation reaction limit, and the first secondary pairs are created by photons with \( \gamma \sim m_e c^2 / \tilde{\gamma}_{\text{min}} \), since they have the shortest mean free paths. However, \( \ell_{\gamma\gamma} \) is large even for these photons, so there is a significant delay between the emission of these photons and the screening of the gap, during which time the primary particles are accelerated to even higher energies, e.g., \( \gamma \gtrsim 10 \tilde{\gamma}_{\text{min}} \). The screening process in the KN regime depends strongly on the acceleration history of the primary particles in the gap. When \( \ell_{\gamma\gamma} \gtrsim r_g \), the gap is global and the electric field grows so quickly that not enough photons with energy \( \gtrsim m_e c^2 / \tilde{\gamma}_{\text{min}} \) are emitted. As a result, the gap can no longer be screened.

![Figure 1](image-url)
We discovered that with increasing $\tilde{B}\tilde{\varepsilon}_{\text{min}}$, numerical simulations become more and more challenging. Figure 5 shows light-curve comparisons between simulations with identical parameters, but different numbers of particles per cell, or different resolutions, at $\tilde{B}\tilde{\varepsilon}_{\text{min}} = 10^4$. With a smaller number of particles per cell or lower resolution, the simulation eventually settles to a quasi-steady state not unlike what was reported by Levinson & Cerutti (2018). However, with more particles per cell or higher resolution, we observe the same kind of quasiperiodic opening and screening of the gap as shown in Section 3.1. Generally we find that when the system settles to a quasi-steady state, not only is the dissipation much less than the peak of the cyclic solution, the average power is also about a factor of $\sim 1.5$ less than the average power of the cyclic solution.

This resolution dependent behavior mainly happens with high $\tilde{B}$ and $\tilde{B}\tilde{\varepsilon}_{\text{min}}$, where large electrostatic oscillations can accelerate particles close to $\gamma \sim 1/\tilde{\varepsilon}_{\text{min}}$, and thus sustain pair creation without forming a time-dependent gap. To minimize this numerical effect, one needs both high spatial resolution and a large number of particles per cell. The spatial grid should ideally resolve the plasma skin depth $\lambda_p$, otherwise the plasma will be heated to Lorentz factors $\gamma_p \sim (\Delta x/\lambda_p)^2$, at which point the hot plasma skin depth can be resolved. If this energy is already capable of producing pairs, we will simply have a hot thermal plasma in pair creation equilibrium. A small number of particles per cell will lead to high discrete particle noises in the simulation, which also translates to larger plasma oscillations, broadening the spread of distribution function in momentum space. If this allows more particles to reach pair creation energies then again a quasi-steady state can be achieved. We believe this is the main reason CY18 and Levinson & Cerutti (2018) found qualitatively different solutions to the same problem. Both were using a similar numerical resolution, but Levinson & Cerutti (2018) were attempting to use parameters corresponding to $\tilde{B}\tilde{\varepsilon}_{\text{min}} \sim 5 \times 10^6$ while we found that $\tilde{B}\tilde{\varepsilon}_{\text{min}} = 10^4$ is already very challenging for the typical numerical resolution used in this kind of simulation. For this

Figure 4 shows the dependence of the gap power on $\tilde{B}\tilde{\varepsilon}_{\text{min}}$ and $\tau_0$ individually. In all the runs shown on the plot, we observe the same time-dependent gap described in Section 3.1, and measure the peak gap power averaged over several cycles. It can be seen that in general the power scales close to $L \sim (\tilde{B}\tilde{\varepsilon}_{\text{min}})^{-1}$ with fixed $\tau_0$. However, we have to note here that at high $\tilde{B}\tilde{\varepsilon}_{\text{min}}$ the measured gap power could be smaller than it should be, due to numerical heating, which will be discussed in Section 3.3, since numerical heating tends to reduce the need for gap acceleration.

Also in Figure 4 one can see that when $\tau_0$ is around 10 the gap power scales weakly with $\tau_0^{1/2}$. However, above $\tau_0 \sim 50$ which is around the transition between Thomson and KN regime the power scales closer to $\tau_0^{0.9}$. This being steeper than the scaling of $\tilde{D}_{\text{max}}$ suggests that the gap size decreases with increasing $\tau_0$ as well.
reason, we also expect that the gap power we see in our highest \( \tilde{B}_0 \tilde{\epsilon}_{\text{min}} \) simulations may not be the most reliable (e.g., the last point in the left panel of Figure 4, where \( \tilde{B}_0 \tilde{\epsilon}_{\text{min}} = 3 \times 10^4 \)), as numerical effects reduce the amount of \( E_{\parallel} \) generated by the system.

4. Implications for M87

M87, as a low-luminosity AGN, is one of the most promising candidates where this process can operate. One can obtain a reasonable estimate for the magnetic field from the estimated jet power, \( B_0 \sim 200 \) G (e.g., Levinson & Rieger 2011; Broderick & Tchekhovskoy 2015), which translates to \( B_0 \sim 1.2 \times 10^{14} \). However, the parameters for the soft photon field in the vicinity of the central black hole is a lot more uncertain. The SED for the radio flux seems to peak around 1.2 meV, which translates to \( \tilde{\epsilon}_{\text{min}} \sim 2 \times 10^{-9} \), but the soft photon spectrum near the horizon may well be very different from what is observed. Assuming all photons are coming from near the horizon, the upper bound for photon energy density is \( u_{s} \sim 0.1 \text{ erg cm}^{-3} \) which implies that \( \tau_0 \sim 5 \times 10^3 \). But since most photons are produced in the disk, \( u_{s} \) near the horizon could be orders of magnitude lower.

In CY18 we found that if we take \( \tau_0 \sim 5 \times 10^3 \) then the gap power in the Thomson regime would be \( L \sim 3 \times 10^{39} \) erg s\(^{-1}\). If we assume much lower fiducial optical depth, e.g., \( \tau_0 \sim 10 \) as in our simulations, and simply scale \( L / L_0 \) according to Figure 4 to a realistic value of \( \tilde{B}_0 \tilde{\epsilon}_{\text{min}} \sim 2.4 \times 10^5 \), we can expect the power dissipated in the gap to be \( L \sim 10^{-3} L_0 \). The jet power of M87 is usually estimated to be \( L_0 \sim 10^{33} \)–\( 10^{34} \) erg s\(^{-1}\), so our estimated gap power should be around \( L_{\text{gap}} \sim 10^{30} \)–\( 10^{31} \) erg s\(^{-1}\), somewhat smaller than, but not entirely inconsistent with the reported power of the VHE flares from the same object (e.g., Abramowski et al. 2012).

We expect the spectrum from the gap to be similar to what is shown in Figure 3, which is a power law extending up to \( \sim 0.1 / \tilde{\epsilon}_{\text{min}} = 5 \times 10^7 \), followed by an exponential cutoff. This translates to energy up to around 25 TeV, also consistent with the observed flare spectrum. The exact slope of the spectrum will likely be subject to further modification due to scattering with the ambient soft photon field and radiation from the secondary pairs.
5. Discussion

We presented results from first-principle 1D GR PIC simulations of the pair discharge process in the vicinity of supermassive black holes. In general we observe a highly time-dependent solution where an electric gap opens quasiperiodically near the null surface. The variability timescale of this activity is usually on the order of several \( r_g/c \), and the outgoing photon spectrum forms a power law with index that depends on the gap activity.

In this work we ignored synchrotron and curvature radiation. In reality, when \( \tau \) is order unity, the photon free path is significant compared to the macroscopic length scale. As they convert to a pair, the particles should have a relatively large pitch angle with respect to the magnetic field. It will be damped quickly through synchrotron radiation, which should dissipate a sizable fraction of the particle energy and produce detectable radiation. The highest energy particles that have \( \gamma \sim 10/\epsilon_{\min} \) can approach the curvature radiation reaction limit, and curvature loss can be a significant energy sink as well. We plan to carry out more complete spectral analysis in the future, including the effect of synchrotron and curvature radiation in this framework, which could produce a unified spectrum from X-ray to VHE gamma-rays out of the pair producing gaps.

In this paper we considered \( \gamma-\gamma \) collision to be the only pair creation mechanism. As pointed out by Petropoulou et al. (2019), when particles are in the deep KN regime, triplet pair production may play a role in the screening of the gap. Due to the highly uncertain nature of the soft photon field in the vicinity of the supermassive black hole of M87, there is a parameter regime where triplet pair production could be the main mechanism that regulates the gap physics. This can potentially further increase the gap luminosity and could be powering the strongest flares that we see up to \( L \sim 10^{42} \text{ erg s}^{-1} \). A radiation module that handles triplet pair creation is already implemented in Aperture and we expect to carry out a separate study for the importance of this process in M87, and whether this produces any observable signatures on the spectrum.

This paper studied the gap microphysics in 1D, whereas in many runs the gap evolves to have a size comparable to \( r_g \), at which point the 1D approximation is not completely valid. Working 2D axisymmetric GR PIC codes now exist (Parfrey et al. 2019). It is an important and interesting question whether the results presented in this paper are robust when the 1D approximation is relaxed. However, we caution any future attempt at solving this problem in 2D full GR that resolution is important. Significant rescaling and some abstraction of the radiative physics must be employed, and extreme care must be taken in order to draw any meaningful comparisons with observations.

We thank Amir Levinson, Frank Rieger, Kouichi Hirotani, Zhen Pan, and Huan Yang for helpful discussion. The code Aperture used in this work can be found at https://github.com/zban007/Aperture3.git. A.C. acknowledges support from NASA grant 80NSSC18K1099. Y.Y. acknowledges support from the Lyman Spitzer, Jr. Postdoctoral Fellowship provided by the Department of Astrophysical Sciences, Princeton University, and a Flatiron Research Fellowship provided by the Flatiron Institute, Simons Foundation. We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research.

Appendix

Appendix Background Force-free Configuration

We consider a split-monopole field configuration in Kerr spacetime. The axisymmetric, steady-state, force-free solution is obtained using a numerical method described in Yuan et al. (2019). Figure A1 shows one example with \( a = 0.99 \), which we used in this paper.

The background solution is specified by the magnetic flux \( \psi \) and two flux functions: \( \Omega(\psi) \)—angular velocity of the field line, and \( B_r(\psi) \)—current flowing along the flux surfaces. The components of the magnetic field can be obtained as

\[
B^\phi = \frac{1}{\sqrt{g}} B_T g_{\phi\psi} g_{\theta\theta}, \quad B^\theta = \frac{\partial \psi}{\sqrt{g}}, \quad B^\phi = -\frac{\partial r \psi}{\sqrt{g}}. \tag{A1}
\]

The background charge and current densities satisfy

\[
j^\phi_\Omega = \Omega \rho_\Omega + \frac{B^\phi}{B^\psi} j^\psi, \tag{A2}
\]

and the corresponding components are

\[
\rho = \alpha J^0 = -\frac{1}{\sqrt{g}} \partial_\theta \left( \frac{\sqrt{-g}}{\Delta \sin^2 \theta} \theta^{AB}(\Omega - \omega) \partial_B \psi \right),
\]

where \( A, B = r, \theta \), \( \omega = \frac{B_T r}{\sqrt{g}} \), \( j^\phi = \alpha j^\phi = \frac{B_T r}{\sqrt{g}}. \tag{A3}\)

ORCID iDs

Alexander Y. Chen @ https://orcid.org/0000-0002-4738-1168
Yajie Yuan @ https://orcid.org/0000-0002-0108-4774
References

Abramowski, A., Acero, F., Aharonian, F., et al. 2012, ApJ, 746, 151
Aharonian, F. A., Barkov, M. V., & Khangulyan, D. 2017, ApJ, 841, 61
Alesksic, J., Ansoldi, S., Antonelli, L. A., et al. 2014, Sci, 346, 1080
Beskin, V. S., Istomin, Y. N., & Parev, V. I. 1992, SvA, 36, 642
Blandford, R. D., & Znajek, R. L. 1977, MNRAS, 179, 433
Blumenthal, G. R., & Gould, R. J. 1970, RvMP, 42, 237
Broderick, A. E., & Tchekhovskoy, A. 2015, ApJ, 809, 97
Chen, A. Y., Yuan, Y., & Yang, H. 2018, ApJL, 863, L31
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, ApJL, 875, L1
Ford, A. L., Keenan, B. D., & Medvedev, M. V. 2018, PhRvD, 98, 063016
Gould, R. J., & Schréder, G. P. 1967, PhRv, 155, 1404
Hirotani, K., & Okamoto, I. 1998, ApJ, 497, 563
Hirotani, K., & Pu, H.-Y. 2016, ApJ, 818, 50
Hirotani, K., Pu, H.-Y., Lin, L. C.-C., et al. 2016, ApJ, 833, 142
Hirotani, K., Pu, H.-Y., Lin, L. C.-C., et al. 2017, ApJ, 845, 77
Jones, F. C. 1968, PhRv, 167, 1159
Katsoulakos, G., & Rieger, F. M. 2018, ApJ, 852, 112
Komissarov, S. S. 2004, MNRAS, 350, 427
Levinson, A., & Rieger, F. 2011, ApJ, 730, 123
Levinson, A., & Segev, N. 2017, PhRvD, 96, 123006
McKinney, J. C., Tchekhovskoy, A., & Blandford, R. D. 2012, MNRAS, 423, 3683
Parfrey, K., Philippov, A., & Cerutti, B. 2019, PhRvL, 122, 035101
Petropoulou, M., Yuan, Y., Chen, A. Y., & Mastichiadis, A. 2019, ApJ, 883, 66
Ptitsyna, K., & Neronov, A. 2016, A&A, 593, A8
Rybicki, G. B., & Lightman, A. P. 1986, Radiative Processes in Astrophysics (New York: Wiley-VCH)
Timokhin, A. N., & Arons, J. 2013, MNRAS, 429, 20
Yuan, Y., Blandford, R. D., & Wilkins, D. R. 2019, MNRAS, 484, 4920