Spectral and imaging properties of Sgr A* from high-resolution 3D GRMHD simulations with radiative cooling


1 Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands
2 Gravitation and Astroparticle Physics Amsterdam (GRAPPA) Institute, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands
3 Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK
4 Institut für Theoretische Physik, Goethe-Universität Frankfurt, Max-von-Laue-Straße 1, D-60438 Frankfurt am Main, Germany
5 Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA
6 Center for Interdisciplinary Exploration & Research in Astrophysics (CIERA), Physics & Astronomy, Northwestern University, Evanston, IL 60202, USA

Accepted 2020 September 28. Received 2020 September 28; in original form 2020 August 17

ABSTRACT

The candidate supermassive black hole in the Galactic Centre, Sagittarius A* (Sgr A*), is known to be fed by a radiatively inefficient accretion flow (RIAF), inferred by its low accretion rate. Consequently, radiative cooling has in general been overlooked in the study of Sgr A*. However, the radiative properties of the plasma in RIAFs are poorly understood. In this work, using full 3D general-relativistic magnetohydrodynamical simulations, we study the impact of radiative cooling on the dynamical evolution of the accreting plasma, presenting spectral energy distributions and synthetic sub-millimetre images generated from the accretion flow around Sgr A*. These simulations solve the approximated equations for radiative cooling processes self-consistently, including synchrotron, bremsstrahlung, and inverse Compton processes. We find that radiative cooling plays an increasingly important role in the dynamics of the accretion flow as the accretion rate increases: the mid-plane density grows and the infalling gas is less turbulent as cooling becomes stronger. The changes in the dynamical evolution become important when the accretion rate is larger than $10^{-8} \, M_\odot \, \text{yr}^{-1}$ ($\gtrsim 10^{-7} \, M_{\text{Edd}}$, where $M_{\text{Edd}}$ is the Eddington accretion rate). The resulting spectra in the cooled models also differ from those in the non-cooled models: the overall flux, including the peak values at the sub-mm and the far-UV, is slightly lower as a consequence of a decrease in the electron temperature. Our results suggest that radiative cooling should be carefully taken into account in modelling Sgr A* and other low-luminosity active galactic nuclei that have a mass accretion rate of $\dot{M} > 10^{-7} \, M_{\text{Edd}}$.

Key words: accretion, accretion discs – black hole physics – MHD – methods: numerical – stars: jets – galaxies: individual: (SgrA*).

1 INTRODUCTION

It is widely believed that most galaxies harbour supermassive black holes (SMBHs) in their galactic centres, with masses ranging from millions to billions of solar masses. Over the past few decades, the black hole (BH) candidate in the centre of the Milky Way, Sagittarius A* (hereafter Sgr A*), has proven an exceptional laboratory for studies of accretion and outflow physics of BHs due to its proximity to Earth. A significant effort has been invested in determining the BH mass and distance for Sgr A* (e.g. Reid 1993; Schödel et al. 2002; Bower et al. 2004; Ghez et al. 2008; Gillessen et al. 2009, 2017; Boelette et al. 2016; Gravity Collaboration 2018; Reid et al. 2019). We adopt the current best-fitting BH mass of $4.1 \pm 0.03 \times 10^6 \, M_\odot$, which was measured by the orbital motion of stars and gas clouds (Gillessen et al. 2017; Gravity Collaboration 2018), and the distance of $8.15 \pm 0.15 \, \text{kpc}$, which was obtained from trigonometric parallaxes and proper motions of massive stars around Sgr A* (Reid et al. 2019). Given the mass and the distance, the angular size of the Schwarzschild radius, $r_s = 2GM_{\text{BH}}/c^2$, is $\approx 10 \, \mu\text{as}$, which subsumes a larger area in the sky than any other known BH, including all stellar-mass BHs.

Recently, mounting attention has been paid to the study of Sgr A* with the advent of the pioneering instruments GRAVITY (Gravity Collaboration 2017, 2018) and the Event Horizon Telescope (EHT; Doelman et al. 2009; Event Horizon Telescope Collaboration 2019), capable of probing 10–30 $\mu$as scales. These instruments allow us to profoundly improve our understanding of the physical processes associated with the accretion and relativistic jet formation in the immediate vicinity of SMBHs and thus demand equal measures of theoretical support and precise modelling of the spectrum generated by the radiation from the accretion flow around Sgr A*.

The mass accretion rate around Sgr A* (in units of $M_\odot \, \text{yr}^{-1}$) is estimated to be in the range of $\sim 10^{-9} < \dot{M} < 10^{-7}$, as constrained by the measured Faraday rotation measure at mm/sub-mm wavelengths (Aitken et al. 2000; Bower et al. 2003; Marrone et al. 2007). Such a low accretion rate favours hot accretion flow models...
for the accretion disc instead of the radiatively efficient, thin disc models (Shakura & Sunyaev 1973). Many theoretical scenarios have been invoked and excluded to account for the nature of accretion and outflows in the hot accretion flow: the standard advection-dominated accretion flow (ADAF; Narayan & Yi 1994; Narayan, Yi & Mahadevan 1995; Narayan et al. 1998) and the Bondi–Hoyle models are ruled out, since these models are expected to yield an accretion rate of $\sim 10^{-5} M_{\odot} \text{yr}^{-1}$, which is two orders of magnitude higher than the measured upper limit (Bower et al. 2003). Yuan, Quataert & Narayan (2003) re-examined the radiatively inefficient accretion flow (RIAF) model for the spectrum of Sgr A* and argued that the presence of outflows within the Bondi radius plays a vital role in reducing the mass accretion rate. Alternatively, the convection-dominated accretion flow (CDAF) and the jet-dominated models are capable of reproducing the spectrum that is consistent with the observed accretion rate (Falcke & Markoff 2000; Quataert & Gruzinov 2000; Markoff, Bower & Falcke 2007). The existence of such outflows is supported by the weak hydrogen-like Fe Kα line around Sgr A* via the X-ray Visionary Program (Wang et al. 2013). In this study, the flat density profile in the spectrum confirmed that $\gtrsim 99$ per cent of the matter initially captured by the SMBH is lost before it reaches the innermost region around Sgr A*, which is consistent with the CDAF model or adiabatic inflow–outflow solution (ADIOS; see Blandford & Begelman 1999; Yuan & Narayan 2014 for the detailed model descriptions) model.

Although semi-analytic models provide an important framework for understanding the nature of the accretion flow around Sgr A*, numerical simulations are required to capture the time-dependent, turbulent evolution of the accretion flow. In particular, a self-consistent magnetohydrodynamical (MHD) description enables us to demonstrate accretion processes induced by the magnetorotational instability (MRI; Balbus & Hawley 1991) without imposing an arbitrary anomalous viscosity. In earlier numerical studies, 3D pseudo-Newtonian MHD simulations were carried out to model the synchrotron radiation from accretion flows (e.g. Goldston, Quataert & Igumenshchev 2005; Ohshima, Kato & Mineshige 2005; Chan et al. 2009; Huang et al. 2009). However, the non-relativistic treatment in the simulations has disadvantages for modelling the synchrotron radiation, since it is mainly emitted in the immediate vicinity of the central BH, where relativistic effects cannot be ignored: shocks develop differently for the relativistic plasma when subject to intense magnetic and gravitational fields (Del Zanna, Bucciantini & Londrillo 2003), and the curvature of space–time becomes significant. Several other works made use of general–relativistic magnetohydrodynamical (GRMHD) simulations for studying the dynamics and spectral properties of Sgr A* in two dimensions (2.5D; e.g. Noble et al. 2007; Mościbrodzka et al. 2009; Hibburn et al. 2012; Dibi et al. 2012; Drappeau et al. 2013; Mościbrodzka & Falcke 2013), or in three dimensions (3D) (e.g. Dexter, Agol & Fragile 2009; Dexter et al. 2010; Dolence et al. 2012; Schcherbakov, Penna & McKinney 2012; Dexter & Fragile 2013; Mościbrodzka et al. 2014; Chael et al. 2018; Davelaar et al. 2018). In general, 2.5D simulations are a reasonable and computationally cheaper option to conduct a parameter study for reproducing the spectrum of Sgr A*, but it is known that simulations with axisymmetric coordinates cannot sustain MRI-driven turbulence, which decays over the local orbital time as a consequence of Cowling’s antidynamo theorem (Cowling 1933, see also Hide & Palmer 1982 for the generalized description). Therefore, full 3D GRMHD simulations are necessary to perform detailed studies of the nature of the accretion flows around Sgr A* and the emitted spectrum. For instance, it was confirmed that thick accretion discs are able to generate and advect large-scale poloidal magnetic flux through dynamo action when resolved properly (Liska, Tchekhovskoy & Quataert 2018a).

The bolometric luminosity of Sgr A* is extremely low, $L_{\text{bol}} \sim 10^{36} \text{erg s}^{-1} \approx 10^{-9} L_{\text{Edd}}$, where $L_{\text{Edd}}$ is the Eddington luminosity. Given such a low luminosity, it has been thought that the radiative cooling losses of Sgr A* are negligible, since the losses are likely not strong enough to have a significant impact on the dynamics of the accretion flow. Based on this argument, all previous works with full 3D GRMHD simulations have ignored the radiative cooling losses for simplicity. Although this assumption may be reasonable, Dibi et al. (2012) argued based on their 2.5D simulations that cooling losses play an increasingly important role for higher accretion rates and possibly alter the dynamics and resulting spectra of Sgr A*, even within the allowed range of accretion rates based on polarization and X-ray studies. One potential impact is that the radiative cooling reduces the gas pressure and the disc vertical scale height, resulting in a decrease in turbulence and a more ordered magnetic field (Fräule & Meier 2009). Moreover, many questions remain unanswered: how do radiative cooling losses affect the turbulence features of the disc, and thus the angular momentum transfer of the accreting plasma? How does radiative cooling together with GR effects result in the observed spectra from Sgr A*? Is radiative cooling indeed negligible for the mass accretion rate range of Sgr A*? Even though the effects of radiative cooling can be minor for the case of Sgr A*, the quantitative evaluation of cooling effects is highly demanded because it must play a greater role for SMBHs with higher mass accretion rates, such as M87.

In this paper, we perform the first full 3D GRMHD simulations, which include cooling losses via bremsstrahlung, thermal synchrotron emission, and inverse Compton scattering. Due to the significant computational expense of the full 3D simulations, we cannot explore the full range of various parameters (e.g. spin, magnetic configuration, electron distribution function, misaligned disc). Instead, we use parameters compatible with earlier studies, assuming a rapidly rotating BH, weak poloidal initial magnetic field, and a fixed temperature ratio between ions and electrons of $T_e/T_i = 3$ (we also carry out additional simulations with different electron temperature prescriptions for comparison). We then examine the effect of radiative cooling on the dynamics of the accretion flow and the resulting spectra and images for different accretion rates within the allowed range.

This paper is structured as follows. In Section 2, we give a technical description of the numerical methods used, including the simulation setup, and the treatment of radiative cooling losses. In Section 3, we describe the results of how cooling losses play a role in changing the dynamical evolution of accreting matter. In Section 4, we discuss the best-fit model for Sgr A*, the effects of cooling on the resulting spectra and sub-mm images, and the variability of multiwavelength spectra. We also compare our 3D work to previous 2.5D work. We summarize our results in Section 5.

2 TECHNICAL DESCRIPTION OF METHOD

All simulations are performed with the H–AMR code (Liska et al. 2019a; Porth et al. 2019), which branched off HARM2D (Gammie, McKinney & Tóth 2003; Noble et al. 2006) in its early days. It is accelerated by graphical processing units and improved with a staggered grid for constrained transport of magnetic fields (Gardiner & Stone 2005) to preserve $\nabla \cdot \mathbf{B} = 0$, more robust inversion (Newman & Hamlin 2014) adaptive mesh refinement (AMR, not utilized in this work), static mesh refinement, and a locally adaptive time-step (see Chatterjee et al. 2019, Appendix A). It adopts a piece-wise
parabolic method (Colella & Woodward 1984) for reconstruction of cell-centred quantities at cell faces, which is third-order accurate, for the spatial reconstruction at cell faces from cell centres, and a second-order time-stepping.

The broad-band spectrum is calculated from the GRMHD output, using the general–relativistic Monte Carlo scheme GRMONTY (Dolence et al. 2009), which includes synchrotron emission and absorption, and inverse Compton scattering for a relativistic thermal Maxwell–Jüttner distribution of electrons. Technically, GRMONTY cannot produce synthetic images but only spectra. Thus, we ray-trace the GRMHD-produced spectra by integrating the general–relativistic radiative transfer (GRRT) equations using the BHOS code (Younsi, Wu & Fuerst 2012; Younsi et al. 2016, 2020) to generate synthetic images at 230 GHz that can help us to infer the expected images of Sgr A* from the upcoming EHT project. In BHOS, the calculation of radiative processes includes synchrotron emission and absorption only, which is sufficient for imaging at the EHT frequency of 230 GHz. Since the sub-mm regime of the spectra is dominated by synchrotron emission, which both codes calculate, we verify consistency in our spectral calculations from both codes by comparing the resulting spectra in the radio to near-infrared (NIR) bands (see Appendix A).

2.1 Numerical grid and floors

For convenience, we adopt Advection–Lorentz units, which absorb a factor of $\sqrt{4\pi}$ for the magnetic field 4-vector, $b^\mu$, so that the magnetic pressure is $P_b \equiv b^\mu b^\mu$. Furthermore, the typical units are used, $G = c = 1$, which sets the length unit to be the gravitational radius, $r_g \equiv GM/c^2$, and the time unit to be the light crossing time, $t_g \equiv GM/c^3$, where $G$, $M$, $c$ are the gravitational constant, BH mass, and the speed of light, respectively. We use a spherical–polar axisymmetric computational grid ($r$, $\theta$, $\phi$) extending from $0.85r_H$ to $250r_g$, where the event horizon radius $r_H \equiv r_g\left(1 + \sqrt{1 - a^2}ight)$.

Here, we set the dimensionless BH spin parameter to $a_* \equiv cJ/GM^2$ in a Kerr–Schl fid foliation, where $J = r_Ha_*$ is the angular momentum at the event horizon. The grid is uniformly spaced with respect to a set of internal coordinates $(x^1, x^2, x^3)$, which can be converted to $(r, \theta, \phi)$, respectively. This conversion leads to a logarithmic spacing in $r$ such that the cells have a higher resolution for smaller values of $r$. The spatial resolution near the event horizon is $\Delta r \approx 0.01r_g$ for the model with the highest resolution. To prevent the aspect ratio of the cells from becoming too large near the polar singularity, we reduce the resolution in $\phi$ direction gradually towards both poles. We use outflow boundary conditions for both inner and outer radial boundaries and reflecting boundary conditions in the $\theta$ direction. Note that the inner boundary is causally disconnected from the flow, as it is located within the event horizon.

It is common for GRMHD simulations to crash if either the density or the internal energy becomes very low, particularly in the funnel region along the polar axes or near the outer radial boundaries. To avoid this, we apply numerical floors for the density and the internal energy (see Appendix B3 of Ressler et al. 2017 for more detailed discussions): a minimum rest mass density, $\rho_{\text{min}} = \max[b^2/20, u_g/150, 10^{-4}(r/r_g)^{-2}]$, and a minimum internal energy density, $u_{\text{g},\text{min}} = \max[b^2/750, 10^{-4}(r/r_g)^{-2}]$, where $b$ and $u_g$ are the co-moving magnetic field strength and the internal energy density, respectively. We normalize the mass density such that the maximum density is $\rho_{\text{max}} = 1$.

2.2 Simulation models

We perform a set of GRMHD simulations, in which the magnetized gas is accreting on to a supermassive and spinning BH. All simulations are initialized with a steady-state hydrostatic torus around a rapidly spinning Kerr BH (Fishbone & Moncrief 1976). We set the spin parameter to $a_* = 0.9375$ for all models. The size of the initial torus is set by the inner edge, $r_{\text{in}} = 6r_g$, and the radius of the pressure maximum, $r_{\text{max}} = 12r_g$. We adopt an ideal gas equation of state,

$$P_g = (\Gamma - 1) u_g, \quad (1)$$

where $P_g$ and $u_g$ are the gas pressure and internal energy, respectively.

We set the adiabatic index to $\Gamma = 5/3$, which assumes the dominance of a non-relativistic plasma in the accretion flow.

As an initial magnetic configuration, we adopt a single loop of weak magnetic field, which is computed from the magnetic vector potential,

$$A_\phi \propto \left\{ \begin{array}{ll} \rho - 0.2, & \text{if } \rho > 0.2, \\ 0, & \text{otherwise.} \end{array} \right. \quad (2)$$

The centre of the loop is at the density maximum, and the loop is fully contained within the initial torus. The initial magnetic field is normalized such that $B_{\text{max}} = P_g/P_B \geq 100$. This normalization ensures that the magnetic pressure is subdominant compared to the gas pressure.

2.3 Radiative cooling

We take into account the radiative cooling self-consistently in our calculation of the gas temperature by including the effects of bremsstrahlung, synchrotron, and the inverse Compton losses. We adopt the equations of Esin et al. (1996) for computing the radiative cooling losses. These formulae have been implemented and tested in previous numerical studies of Sgr A* (Fragile & Meier 2009; Dibi et al. 2012; Straub et al. 2012; Drappeau et al. 2013).

The total cooling rate for an optically thin gas is computed from the cooling function,

$$q_{\text{th}} = \eta_{\text{bc}} c q_{\text{bc}} + \eta_{\text{sc}} c q_{\text{sc}} + q_{s} + q_{a},$$ \quad (3)

where $q_{\text{bc}}$ and $q_{s}$ are the bremsstrahlung and synchrotron cooling rates, respectively, and $\eta_{\text{bc}} c$ and $\eta_{\text{sc}} c$ are the Compton enhancement factors, which are the average energy gain of the photon in an assumption of single scattering (Esin et al. 1996). We note that the Compton enhancement of the bremsstrahlung is negligible as synchrotron is dominant at the temperature where the Comptonization becomes important.

While the whole system is generally optically thin, we use the following generalized cooling formula, from Narayan & Yi (1995) and Esin et al. (1996), to reproduce the equilibrium solution corresponding to the optically thick disc (Shakura & Sunyaev 1973):

$$q_{\text{e}} = \frac{4\sigma_T T_e^4}{1.5r_T^{\frac{4}{3}}} + \frac{1}{\sqrt{3} + \tau_{\text{bs}}},$$ \quad (4)

where $\sigma_T$ and $T_e$ are the Thomson cross-section and the electron temperature, respectively, and the local temperature scale height $H_T$ is computed from

$$H_T = \frac{T_e}{|\nabla(T_e^4)|}.$$ \quad (5)
The scale heights are locally calculated such that $T^4$ drops off by a factor of $1/e$, which was adopted in Fragile & Meier (2009) as a suitable and robust treatment in multidimensional simulations.

The total optical depth of the disc is calculated by $\tau = \tau_{\text{opt}} + \tau_{\text{abs}}$, where $\tau_{\text{opt}} = 2n_e \sigma_T H$ is the Thomson optical depth in the vertical direction and $\tau_{\text{abs}}$ is the optical depth for absorption, which is expressed as

$$\tau_{\text{abs}} = H_T \frac{q_{\text{fin}}}{{4\pi} T_e^4}. \quad (6)$$

For a small optical depth, equation (4) reduces to equation (3), while, in the optically thick limit ($\tau \gg 1$), it gives $q^- = 8\pi \sigma_T T_e^4/3H_T^2\tau$, which is the appropriate blackbody limit. Therefore, the formula provides an approximate interpolation between the optically thin and thick limits.

At low temperatures ($T_e \lesssim 6 \times 10^9 K$) or the outer torus regions, the emission is dominated by bremsstrahlung (Straub et al. 2012). The bremsstrahlung cooling rate is computed by the interactions of pairs among electrons (e$^-$), positrons (e$^+$), and ions (i). Since the cooling processes of e$^-$i and e$^+i$ are identical, and the same is true for e$^-e^+$ and e$^+e^-$, the cooling rate can be written as,

$$q^e = q^e_+ + q^e_-$,$$$$

where $q^e_+$, $q^e_-$, and $q^e_m$ are the radiative cooling through electron-positron–ion (e$^+i$), electron-positron–electron-positron (e$^+e^-$), and positron–electron (e$^-e^-$) interactions, respectively (Esin et al. 1996).

However, for most regions of inner hot accretion flows, the synchrotron emission dominates the losses as the electrons are relativistic due to the high electron temperature. The synchrotron cooling occurs through both optically thick and thin emission: below some critical frequency $\nu_c$, the emission is completely self-absorbed, and thus the volume emissivity can be approximated by the Rayleigh–Jeans blackbody emission. For frequencies above $\nu_c$, the emission is dominated by bremsstrahlung (Straub et al. 2012).

The bremsstrahlung cooling rate is computed by the interactions of pairs among electrons (e$^-$), positrons (e$^+$), and ions (i). Since the cooling processes of e$^-i$ and e$^+i$ are identical, and the same is true for e$^-e^+$ and e$^+e^-$, the cooling rate can be written as,

$$q^e = \frac{2\pi k_B T_e}{H_T e^4} \int_0^{\nu_c} \nu^2 d\nu + \int_{\nu_c}^{\infty} \epsilon_s(\nu) d\nu, \quad (8)$$

where $k_B$ and $c$ are the Boltzmann constant and the speed of light, respectively, and the synchrotron emissivity $\epsilon_s(\nu)$ is calculated as (see Pacholczyk 1970),

$$\epsilon_s(\nu) = \frac{e^2}{c^2} \pi n_e (n_e + n_i) \int I(\nu) d\nu,$$$$

where $K_2$ is the modified Bessel function of the second kind,

$$x_M = \frac{2\nu}{3\sqrt{\pi} \theta_e^2}, \quad v_0 = \frac{\epsilon B}{2\pi m_e c}, \quad (10)$$

and $\theta_e \equiv k_B T_e/m_e c^2$ is the dimensionless electron temperature. The dimensionless spectrum $I(x_M)$, which is averaged over the angle between the velocity vector of the electron and the direction of the local magnetic field, is fitted by the function (Mahadevan, Narayan & Yi 1996),

$$I(x_M) = \frac{4.0505}{x_{\text{M}}^{1/6}} \left(1 + \frac{0.4}{x_{\text{M}}^{1/4}} + \frac{0.5316}{x_{\text{M}}^{1/2}} \right) \exp \left(-1.8899 x_{\text{M}}^{1/3} \right). \quad (11)$$

Fragile & Meier (2009) found that the Bessel function $K_2$ in equation (9) causes errors for the low-temperature flows ($T_e < 10^8 K$) due to the mismatch of the normalization factor between the Bessel function and the spectrum $I(x_M)$. Following their suggested modification, we replace $K_2(1/\theta_e)$ by $2\Theta_e^2$, thereby assuming the same high-temperature limit.

We numerically compute $\nu_c$ in equation (8) by equating the optically thin and thick volume emissivities at $\nu_c$,

$$\epsilon_s(\nu_c) = \frac{e^2}{c^2} \pi n_e (n_e + n_i) I(\nu_c) = \frac{2\pi k_B T_e}{H_T e^4} \nu_c^2 \epsilon_s(\nu_c), \quad (12)$$

3 RESULTS

We initialize our fiducial model by following the ‘best-bet’ model, that is widely agreed by previous 2.5D parameter surveys (e.g. Mościbrodzka et al. 2009; Dibi et al. 2012; Drappeau et al. 2013), which have $T_e/T_v = 3$ and $a_* = 0.9375$ (see Table 1).

3.1 General evolution

Our simulations start with the initial torus in hydrostatic equilibrium (Fishbone & Moncrief 1976). As the turbulence triggered by tangled magnetic fields transports angular momentum outward, the gas flows towards the central BH generating a thick disc, akin to an RIAF. The cooled models require a specific density unit to be pre-set in order to achieve the designated mass accretion rate in the simulations (see Table 1). However, the non-cooled models are scale-free and therefore scaled by the corresponding density unit (in GRRT post-processing) to enable comparison with the cooled models.

Fig. 1 shows the density contour map overlaid with the magnetic field structure, at which the data are averaged for 5000–6000 $t_g$. The overall evolution of the accreting hot accretion flow is similar between the non-cooled and cooled runs when the mass accretion rate is smaller than $10^{-8} \ M_\odot \ yr^{-1}$. However, the effect of cooling becomes increasingly important and shows visible differences in model C3D10RM, where the target mass accretion rate is $10^{-7} \ M_\odot \ yr^{-1}$ (the estimated value is $8.22 \times 10^{-8} \ M_\odot \ yr^{-1}$): when radiative cooling is on, the density increases significantly in the mid-plane and the magnetic field within the disc is less turbulent. This is because cooling reduces the gas pressure and the corresponding scale height of the accretion flow, thus increasing the dominance of magnetic fields: the plasma beta, $\beta_{magn} = \rho_c^2/\rho_g$, decreases due to the reduced gas pressure and the compressed volume. Such highly magnetized plasma tends to be stable against the MRI, and thus the magnetic field within the disc becomes less turbulent.

Radiative cooling enhances the mid-plane density for two main reasons: at first, it is clear that cooling reduces the gas pressure as thermal energy is radiated away. Secondly, the relatively ordered magnetic field impedes angular momentum transport through MRI, so accretion slows down and piles up where the MRI is less efficient. Fig. 2 shows the averaged density profile along the disc over the time interval between 6000 $t_g$ and 8000 $t_g$ by the formula:

$$\rho(\rho)(\phi) := \frac{\int_0^{2\pi} \int_0^{\sqrt{\rho}} \rho^2 \sqrt{-g} d\theta d\phi}{\int_0^{2\pi} \int_0^{\sqrt{\rho}} \sqrt{-g} d\theta d\phi},$$

where $\rho'$ is the time-averaged density. In the figure, the density enhancement is apparent as a consequence of cooling. For models with higher cooling, the relative increase in the density compared to non-cooled runs is larger and occurs over a broader range in radius. More importantly, the peak of the averaged density is located at larger distances and with stronger cooling, which is not surprising

$^2$The orbital time-scale at the pressure maximum, $r_{\text{max}} = 12 r_g$, is $t_{\text{orb}} \sim 260 t_g$ in our simulations, where $t_g \sim 20$ s for Sgr A$^\ast$. 

MNRA 499, 3178–3192 (2020)
A single time-step ($t$) is lines denote the jet boundary, defined as where the magnetization parameter is maximal within the ‘funnel’ over the pole, which is which decreases with increasing distance from the BH. The electron temperature is maximal within the ‘funnel’ over the pole, which is $\sim 10^{12}$ K in the mid-plane. Despite the high temperature in the funnel, it typically produces negligible emission as a consequence of the extremely low densities in this region. Note that since we assume a relativistic thermal Maxwell–Jüttner distribution for the radiative processes, the question still remains as to which fraction of non-thermal electrons can be generated within the plasma by steepening of MHD waves or inducing shocks/turbulence through mechanisms akin to magnetic reconnection, and thus how it contributes to the radio emission (Yuan et al. 2003; Davelaar et al. 2018).

Radiation processes operate predominantly in hot accretion flows. To obtain insight into the physical properties of the flows, we show contour maps slices of density, $B^2$, and electron temperature at a single time-step ($t = 5000\, r_g$) for the highest resolution model (Fig. 3). For the hot accreting plasma, it is evident that the typical density is $\sim 10^{-16}$ g cm$^{-3}$, which corresponds to $N_e \approx 6 \times 10^7$ cm$^{-3}$ in fully ionized plasma, where $N_e$ is the electron number density. The typical strength of magnetic fields is $B \approx 30$ G near the event horizon, which decreases with increasing distance from the BH. The electron temperature is maximal within the ‘funnel’ over the pole, which is up to $\sim 10^{13}$ K, and is $\sim 10^{12}$ K in the mid-plane. Despite the high density in the funnel, it typically produces negligible emission as a consequence of the extremely low densities in this region. Note that since we assume a relativistic thermal Maxwell–Jüttner distribution for the radiative processes, the question still remains as to which fraction of non-thermal electrons can be generated within the plasma by steepening of MHD waves or inducing shocks/turbulence through mechanisms akin to magnetic reconnection, and thus how it contributes to the radio emission (Yuan et al. 2003; Davelaar et al. 2018).

Given the electron temperature range of $10^{10}–10^{12}$ K in the accretion flow, as seen in Fig. 4, bremsstrahlung cooling is relatively weak. Note that in our simulations, the Comptonization of bremsstrahlung is neglected as it is never of importance compared to synchrotron emission over the temperature range of interest. On the other hand, synchrotron cooling with Compton enhancement is dominant in the mid-plane near the BH: optically thin synchrotron radiation at $r \lesssim 5\, r_g$ is responsible for the sub-mm peak in the spectral energy distributions (SEDs), which lies within $10^{11}$ Hz $< \nu < 10^{14}$ Hz (see S4.1.1). Inverse-Compton scattering of synchrotron photons is active at $r \lesssim 8\, r_g$. The emission region and the relative strength of radiative processes are consistent with the previous study (Moscibrodzka et al. 2007), which computed the radiative properties around a hot accretion disc considering the balance of cooling and heating. The
Figure 3. Colour contour maps of the mass density, the magnetic field strength, and the electron temperature for the model C3D1RH. The figure presents a single time slice at 5000 $t_g$.

mean electron temperature in the accretion flow at $r \lesssim 8 r_g$ is $\langle T_e \rangle \approx 3 \times 10^{11}$ K, and therefore the average increase of energy in a single scattering can be approximated as $A = 1 + 4\Theta_e + 16\Theta_e^2 \approx 4 \times 10^4$ (see Esin et al. 1996). As a result, the frequencies of these scattered photons are shifted to the range of 0.02–20 keV, which is consistent with the observed emission in X-rays (0.5–8 keV) (Baganoff et al. 2001, 2003). The quiescent X-ray emission of Sgr A* is extended, with an intrinsic size of $\sim 1.4''$ (Baganoff et al. 2003), which is coincident with the Bondi accretion radius calculated from the measured BH mass and ambient temperature (Yuan et al. 2003). It is known that $\sim 90$ per cent of the total X-ray emission originates from the outer part of the disc (Neilsen et al. 2013) and is dominated by bremsstrahlung, which is beyond the scope of this work: we calculate the emission within $r = 20 r_g$, where the X-ray emission is predominantly produced by the synchrotron self-Compton process. We will further discuss the spectral properties of the X-ray emission in Section 4.1.

3.2 Mass accretion rates

The mass accretion rate is a critical factor in determining the radiation fluxes. In previous works that omit radiative cooling, the simulations are scale-free and must therefore be scaled with an arbitrary density unit during the GRRT post-processing to achieve the desired accretion rate. The scaled variables are used to calculate synthetic spectra, which match with observations (e.g. Dexter et al. 2010; Shcherbakov et al. 2012; Mościobrodzka et al. 2014). Our radiatively cooled simulations, however, are not scale-free: the calculation of cooling rates requires specifying variables in physical units.

Fig. 5 shows the mass accretion rates over time until $t = 10,000 t_g$, which is calculated by

$$M(r) = \int_0^{2\pi} \int_0^{\pi} \rho u' \sqrt{-g} d\theta d\phi,$$

(14)

where $u'$ is the radial component of the 4-velocity. As seen in the figure, the mass accretion rate converges after 3000 $t_g$, which corresponds to $\sim 10$ orbital time-scales at the pressure maximum. This is the case for all models except the low-resolution runs (C3D1RL). The convergence of the accretion rate allows us to study the statistical properties over a longer time period. In contrast to the 3D runs, it is known that previous 2.5D simulations fail to reach the steady state of the accretion as a consequence of the antidynamo theorem (Hide & Palmer 1982). We further compare the results between 2.5D and 3D runs in Section 4.3.

In the lower panel of Fig. 5, we show how the angular momentum transport through the turbulence of the accretion flow can be affected by resolution effects. To examine if the MRI is resolved properly, we calculate the ‘MRI quality factors’ (i.e. Q-factors), which are defined as the number of cells available for resolving the fastest-growing MRI mode in each direction. The Q-factor of the lowest resolution case (C3D1RL) is $\sim 3$, which is below the nominal Q value of 10–20 for capturing the saturation level of the MRI (Hawley, Guan & Krolik 2011). Thus, it is obvious that the mass accretion rate in the model C3D1RL drops significantly after 2000 $t_g$ since the low-resolution run fails to resolve MRI-driven turbulence. However, for the intermediate (C3D1RM) and highest (C3D1RH) resolution cases, the Q-factors are 12 and 20, respectively, which are large enough to sustain the MRI-driven turbulence. These Q-factors are
Figure 4. Colour contour maps of the radiative cooling rates: bremsstrahlung (left), synchrotron (middle), and inverse Compton scattering (right) for the model C3D1RH. The figure presents a single time slice at 5000 \( r_g \). Radiative cooling is calculated only when \( B^2/\rho < 1 \) (see the white area, where the cooling process is off).

Figure 5. The mass accretion rate at the event horizon as a function of time. Upper panel: The solid and dashed lines represent the results from models with radiative cooling and models without radiative cooling (NC3RM), respectively. The accretion rate from the NC3RM model is scaled by the density unit that corresponds to the each cooled counterpart. Lower panel: The accretion rate from the simulations at the different resolutions: C3D1RH (high, dot–dashed), C3D1RM (intermediate, solid), and C3D1RL (low, dotted).

indicative of a criterion above which the simulations satisfactorily resolve the MRI but cannot be used for the analysis of turbulent features in the flow. We will further discuss the disc properties in Section 3.3.

We carry out multiple simulations with different density units to target Sgr A\(^*\) mass accretion rates (in units of \( M_\odot \text{yr}^{-1} \)) of \( 10^{-9}, 10^{-8}, 10^{-7}, \) and \( 10^{-6} \). The largest accretion rate among these target values is beyond the observed range around Sgr A\(^*\) (\( 2 \times 10^{-9} M_\odot \text{yr}^{-1} < M < 2 \times 10^{-7} M_\odot \text{yr}^{-1} \)), but this model is included to compare the results from other simulations with the case of an extremely high accretion rate. In the upper panel of Fig. 5, the solid and dashed lines represent the mass accretion rates, which are calculated from the cooled and non-cooled models, respectively. The non-cooled model is re-scaled by the same density unit for each cooled model. For the model with strong cooling (C3D100RM), it is clear that the overall accretion rate is smaller by a factor of 2, compared to the non-cooled case with the same density unit. However, the models with weak cooling (C3D01RM and C3D1RM) show no significant differences in the accretion rate between the cooled and non-cooled models, which is surprising since the cooled model is expected to lose less angular momentum compared to the non-cooled model. For model C3D10RM, the accretion rate is even slightly higher than in the non-cooled model. The reason for this is the enhanced density in the vicinity of the event horizon playing a role in increasing the accretion rate (see Fig. 2), which compensates for the weak angular momentum transport.
3.3 Disc properties

The direct impact of radiative cooling on the accretion flow can be examined through the disc scale height: cooling decreases the gas pressure and thus renders the disc thinner (Fig. 1). To examine the scale height quantitatively, we compute the formula in Noble, Krolik & Hawley (2010) and Porth et al. (2019), which is expressed as

\[
[H/R](r) := \frac{\int_0^{2\pi} \int_0^{\pi} \rho_0 \sqrt{-g} \left| \theta - \pi/2 \right| d\theta d\phi}{\int_0^{2\pi} \int_0^{r_{\text{max}}} \rho_0 \sqrt{-g} \phi d\phi d\theta}.
\]  

(15)

Fig. 6 shows a clear trend that as cooling becomes stronger, the disc scale height becomes thinner. While the disc swells up rapidly at \( r \lesssim 10r_g \) in the case of weak cooling (C3D1RM), the increase of \( H/R \) is more gradual in the case of stronger cooling (C3D100RM). Except for the case with extremely strong cooling (C3D100RM), the disc scale heights lie within \( H/R = 0.24 - 0.28 \) for radii within 20 \( r_g < r < 50r_g \). As angular momentum is transported outward by the MRI, the gas flows inward and the disc undergoes viscous spreading outwards. Since radiative cooling reduces the MRI turbulence via the enhanced magnetic field strength (see Fig. 1), less spreading of the disc is expected when the cooling becomes stronger. For a more quantitative perspective, we compute the rest-frame density-weighted radius, \( r_d \) [referring to the formula in Porth et al. (2019)], which is expressed as

\[
(r_d)(t) := \frac{\int_0^{2\pi} \int_0^{r_{\text{max}}} \int_0^{r_{\text{max}}} \rho_0 \sqrt{-g} \phi d\phi d\theta d\phi}{\int_0^{2\pi} \int_0^{r_{\text{max}}} \rho_0 \sqrt{-g} \phi d\phi d\theta}.
\]  

(16)

where we set the outer radius of integration to \( r_{\text{max}} = 50r_g \). As seen in Fig. 7, the disc spreading is not distinguishable for models C3D1RM and NC3RM, implying that radiative cooling is not strong enough to affect disc spreading for accretion rates up to \( M = 10^{-9} M_\odot \text{yr}^{-1} \). However, it is apparent that the disc size decreases significantly when the accretion rate is higher than this value. Since current observations of Sgr A* indicate that the accretion rate can reach up to \( M_{\text{Sgr A* max}} = 2 \times 10^{-7} M_\odot \text{yr}^{-1} \) (Marrone et al. 2007), the effects of cooling on the dynamics of the accretion flow should be taken into account even within the range of observationally inferred accretion rates in Sgr A*. The model with the strongest cooling (C3D100RM) shows little spreading over the entire simulation time. Although cooling hinders angular momentum transport as discussed above, the results of C3D100RM may be too dramatic to be considered physically realistic. We found that the MRI Q-factor is reduced to \( 5 - 8 \) for model C3D100RM, as the Alfvén velocity decreases with increasing density in the mid-plane due to the stronger cooling. This range of the Q-factor lies below the criterion for sufficiently capturing the MRI saturation, thereby the significant changes seen in model C3D100RM may be partially caused by the failure to adequately resolve the MRI. Evidently, the resolution also affects disc spreading: as discussed above, the simulations with lower resolutions cannot capture the MRI sufficiently, resulting in suppressed disc spreading (see the dashed blue line; C3D1RL).

4 DISCUSSION

4.1 Radiative properties of Sgr A*

4.1.1 Spectral energy distribution

Once we consider radiative cooling, we can no longer scale the GRMHD data to fit the observed flux. Hence, we choose the best-fitting data that produce the flux lying within the observed ranges at 230 GHz (3 Jy–4.2 Jy; see observations compiled within Connors et al. 2017), but note that this is not a statistical fit. To compare the results between the cooled data and the non-cooled data, we scale the latter with the same mass density unit for each of their cooled counterparts. As seen in Fig. 8, with our fixed value of \( a_\odot = 9375 \) and \( T/T_e = 3 \), the model C3D1RM, at which the target mass accretion rate is \( M = 10^{-9} M_\odot \text{yr}^{-1} \), is reasonably consistent with the observations. The overall shapes of the light curves between the cooled (C3D1RM and non-cooled (NC3RM) models are similar to each other; however, the average fluxes in the non-cooled model are slightly higher than in the cooled model. The average fluxes at 230 GHz, which are calculated for the time after 5000 \( t_g \), are \( 3.63 \pm 0.41 \) Jy and \( 3.8 \pm 0.35 \) Jy for the cooled and non-cooled models, respectively.

We calculate the spectra from the GRMHD simulation data using the Monte Carlo radiative transport code GRMONTY (Dolence et al. 2009), which computes synchrotron emission and absorption, and inverse Compton scattering in full general relativity. Fig. 9 shows the SEDs for the cooled model (C3D1RM) and the non-cooled model (NC3RM) with the same density scale. The SEDs have two peaks: the
Variability studies of Chandra observations showed that ~10 per cent of the total quiescent X-ray emission likely originates from the inner accretion flow (Neilsen et al. 2013; Wang et al. 2013). This indicates that the models can be ruled out in our simulations if they produce X-ray luminosities exceeding $L_X \approx 2.4 \times 10^{32} \text{erg s}^{-1}$. We find that for models with the constant ratio of $T/T_e = 3$, X-ray emission is too strong for both the non-cooled and cooled models at most inclination angles. This implies that the electron temperature should be lower than the value determined by $T/T_e = 3$. In Fig. 10, we compare the post-processed SEDs from the GRMHD data, which are simulated with a different electron temperature prescription. The model with the increased temperature ratio (C3D1RMFT20; $T/T_e = 20$) indeed reduces the X-ray emission to below the observed level. However, it is still problematic because its NIR emission is significantly dimmer than the observed values. This may be attributed to the lack of non-thermal electrons in our simulation. Alternatively, a better-fitting model can be obtained by adopting an electron temperature prescription that depends on the plasma magnetization (Mościbrodzka, Falcke & Shiokawa 2016; Mościbrodzka et al. 2017), which is expressed as

$$T_e = \frac{R_i + \beta^2 R_h}{1 + \beta^2}$$

where $\beta \equiv P_{\text{pol}}/P_{\text{mag}}$, and $R_i$ and $R_h$ are free parameters, which control the dominance of emission depending on the magnetic field strength. The temperature ratio converges into $R_h$ and $R_i$ values at the disc ($\beta \gg 1$) and the jet ($\beta \ll 1$), respectively. We carry out the simulation with one set of $R_i = 1$ and $R_h = 20$, and the resulting spectrum is in good agreement with the observed data (shaded blue line in Fig. 10), except for the mismatch of the NIR power-law slope: it reproduces the observed NIR emission while keeping the X-ray emission within the observed maximum limit in the quiescent state. This reinforces the point that SEDs calculated from GRMHD data can be sensitive to the electron temperature prescription, as was investigated recently by Anantua, Ressler & Quataert (2020) with a wide parameter space in their 'critical beta' electron temperature model and equipartition-based constant electron beta/magnetic bias models.

4.1.2 Synthetic images at 230 GHz

To compute the synthetic mm (230 GHz) images, we use the GRRT code bhoss (Younsi et al. 2012, 2016; Younsi et al. 2020; Younsi et al., in preparation), in which the radiation processes include synchrotron emission and absorption. The inclination angle of both the BH spin axis and the observer line of sight. In general, the fluxes slightly increase with increasing inclination angle. This is mainly due to the orbiting plasma that is approaching the observer and the emission being more strongly Doppler boosted at higher inclination angles (i.e. close to edge-on).

The SED in the cooled model (C3D01RM) differs slightly from the non-cooled model (NC3RM): for the cooled model, the overall flux, including the peak value at the sub-mm bump, is slightly lower than for the non-cooled model. The NIR emission originates from the innermost regions ($2r_g < r < 6r_g$; see Mościbrodzka et al. 2009), where the gas temperature and the magnetic field intensity are high. The relatively weak NIR emission in the cooled model is indicative of the lower gas temperature due to the inclusion of radiative cooling. The peak of the far-UV flux is also slightly higher in the non-cooled model than in the cooled model, mainly due to the higher flux of the seed photons over the NIR band, and the peak frequency in the non-cooled model is ~3.5 times larger than in the cooled model. This is because the average increase of energy in the scattering is formulated to $\Delta = 1 + 4\theta_X + 16\theta_Y^2$ (Esin et al. 1996), implying that the higher temperature in the non-cooled model leads to upscattering of photons into the higher energy range. We note that these differences in the SED arise from our adoption of identical density unit values in the GRRT post-processing for both the cooled and non-cooled models. While the resulting SEDs lie within the observational constraint at 230 GHz (see Fig. 9), if the non-cooled model is normalized in the GRRT calculation by decreasing the density unit to match the flux of the cooled model at 230 GHz, the differences become less significant. The adjusted density unit to achieve this matching of the 230-GHz fluxes is 0.84 times (i.e. smaller than) the cooled model’s density unit value, and thus the estimated accretion rate is also smaller in the non-cooled model.

Figure 8. The synchrotron light curves at 230 GHz for the cooled model (C3D01RM; blue colour) and the non-cooled model (NC3RM; orange colour), which are calculated using BHoss (Younsi et al. 2020; Younsi et al., in preparation). The grey shaded region represents the flux range consistent with observations compiled by Connors et al. (2017). The synthetic spectra (Fig. 9) are calculated after 5500 $t_g$ (vertical dotted line), where the synchrotron flux lies within the observed range for both the cooled and non-cooled models.
Spectra of Sgr A* from 3D radiative GRMHD

Figure 9. Spectral energy distributions of Sgr A*, which are calculated from the simulation results of C3D01RM (cooled, blue) and NC3RM (non-cooled, red). The GRMHD data are averaged over the time interval $t_g$ 6000–10000 $t_g$ (see Fig. 8 for the light curve within this interval). The solid lines represent the mean value of the spectrum, and the shaded regions represent the variation of the spectrum during the time period. Observational data points are taken from Melia & Falcke (2001) and Schödel et al. (2011) at the upper limit of near-to-mid IR band, Connors et al. (2017) in the sub-mm band, Bower et al. (2019) at terahertz frequencies (233, 678, and 870 GHz), and Baganoff et al. (2001) and Baganoff et al. (2003) for X-rays (2–10 keV). The X-ray flux in the ray-traced GRMHD data should be below $\sim 10^{-7}$ erg s$^{-1}$ since most of the X-rays should be emitted from the outer disc via bremsstrahlung (Neilsen et al. 2013; Wang et al. 2013), which is not included in this calculation. The different panels represent the results with different inclination angles: $45^\circ$–$60^\circ$ (left), $60^\circ$–$75^\circ$ (middle), and $75^\circ$–$90^\circ$ (right).

Figure 10. Spectral energy distribution of Sgr A*, calculated using GRMHD results with different electron temperature prescriptions: $T_i/T_e = 3$ (C3D01RM, green line), $T_i/T_e = 20$ (C3D1RMFT20, red line), and $T_i/T_e$ depending on the plasma magnetization (C3D1RMRh20, blue line). The inclination angle is $60^\circ$–$75^\circ$.

accretion discs in a different work (Chatterjee et al. 2020). Fig. 11 shows the synthetic images at 230 GHz for a single time snapshot at 8000 $t_g$, which shows detailed turbulent substructure. The ring-like structure is produced by the gravitational lensing effect that magnifies the emissions from accretion flow, and the bright patches in the left side are the result of Doppler-boosted emissions from the approaching side of the disc approaching the observer. These patches are brighter in the non-cooled model than in the cooled model due to the relatively higher gas temperature near the BH.

In Fig. 12, we take the blurred images, which are averaged over the time interval between 8000 and 9500 $t_g$, as an appropriate proxy for the EHT image. The time interval of 1500 $t_g$ corresponds to $\sim 8.4$ h, given the BH mass is $\sim 4.1 \times 10^6 M_\odot$ yr$^{-1}$ (Gravity Collaboration 2018). The blurred images were obtained through convolution with a Gaussian filter, at which the full width at half-maximum (FWHM) is 20 $\mu$m. As seen in the figure, the emission is dominated by the left side of the disc, which has a symmetric crescent shape for all models. Such a crescent shape is the blurred region of the aforementioned hot patches, which are produced by Doppler beaming. The intensity contrast within the crescent in the non-cooled model is larger compared to the cooled model, as expected from the higher temperature in the regions corresponding to the hot patches.

In general, the BH shadow is clearly visible for lower inclination angles ($\theta_0 < 70^\circ$) but becomes less visible for the edge-on images.

4.2 Variability

The dynamical environment around Sgr A* drives flares through various mechanisms: sudden electron heating by magnetic re-connection, star-disc interactions, stochastic acceleration, gravitational lensing of ‘hotspots’ in the accretion flow, and sudden increases in the mass accretion rate due to the infall of clumps of material (e.g. Markoff et al. 2001; Nayakshin, Cuadra & Sunyaev 2004; Yuan, Quataert & Narayan 2004; Trippe et al. 2007; Hamaus et al. 2009; Gravity Collaboration 2018).

As seen in Fig. 13, both the cooled (C3D1RMRh20) and non-cooled (NC3RM) models are highly variable in the different multi-wavelength bands. The NIR light curve behaves similarly to the X-ray light curve, and the eruption events in both wavebands are roughly correlated with a pronounced rise in the mass accretion rate. Fig. 14...
shows the 230-GHz images for the cooled model C3D1RM during the NIR quiescent state (orange dotted line in Fig. 13) and a flaring state (blue dotted line). In this figure, the optical depth of the accretion flow increases during the flaring event, in accordance with the increase in the mass accretion rate. Fig. 15 shows the corresponding SEDs and clearly illustrates the overall rise in flux across all frequencies. However, the X-ray flux is lower than the quiescent limit even during the flaring event, while the NIR flux exceeds the quiescent flux by a factor of \( \sim 4 \). Hence, an increase in the accretion rate can probably trigger an NIR flare without a clearly detectable X-ray flare, which can account for why some NIR flaring events do not exhibit simultaneous X-ray flares (Hornstein et al. 2007). This can be seen for both the cooled and non-cooled models, implying that radiative cooling perhaps plays little role in producing flares. However, it is expected that such cooling shortens the duration of the flaring events, which its importance can be significant, in tandem with electron heating (Leichtnam et al., in preparation).

However, the maximum peak of NIR emission is \( \sim 0.5 \) mJy, which is an order of magnitude lower than the observed flaring flux (Dodds-Eden et al. 2011). Moreover, the X-ray luminosity lies below the value of \( 2 \times 10^{34} \) erg s\(^{-1}\) for the entire time period, which is still identified as an X-ray ‘quiescent’ state: the observed luminosities of the bright X-ray flares are \( > 10^{35} \) erg s\(^{-1}\) (Haggard et al. 2019). One possible reason is that our assumption of purely thermal electrons is not sufficient to produce NIR flares, since non-thermal electrons can be produced near the BH via relativistic magnetic re-connection (Werner et al. 2016). The contribution of non-thermal electrons to flaring will be discussed in an upcoming paper (Chatterjee et al., in preparation). As an alternative solution, Dexter et al. (2020) suggested that saturation of magnetic flux can trigger the flaring events in a magnetically arrested disc.

### 4.3 Comparison with previous works: 2.5D versus 3D

Recent axisymmetric 2.5D GRMHD simulations have explored the effects of radiative cooling on the dynamical evolution of hot accretion flows around Sgr A* (e.g. Fragile & Meier 2009; Dibi et al. 2012; Straub et al. 2012; Drappeau et al. 2013). However, it is known that MRI-driven turbulence is not sustainable in the axisymmetric 2.5D simulations, which decays over the local orbital time as a consequence of the Cowling’s antidynamo theorem (Hide & Palmer 1982). As seen in Fig. 16, the mass accretion rate in the 2.5D run never reaches a steady state and significantly decreases after 1000 \( t_g \) due to the lack of angular momentum transport via the MRI, while the mass accretion rate in the 3D run reaches a quasi-stationary state. Therefore, the 2.5D, axisymmetric approximation does not allow running over a long simulation time, and instead it requires choosing the data at a certain period of time before the MRI decays dramatically, or including an artificial magnetic dynamo term in the induction equation (Sadowski et al. 2015).

Nevertheless, in many respects our 3D results agree with the previous 2.5D results (Dibi et al. 2012) in that radiative cooling plays an increasingly significant role with increasing mass accretion rate and its impact becomes important above a mass accretion rate of \( \dot{M} > 10^{-8} M_\odot \) yr\(^{-1}\). The best-fitting Sgr A* model with the constant temperature ratio of \( T_i/T_e = 3 \) in our work requires the mass accretion rate of \( \sim 10^{-9} M_\odot \) yr\(^{-1}\), which is similar to the results from the previous 2.5D results (Mościbrodzka et al. 2009; Dibi et al. 2012).
Spectra of Sgr A* from 3D radiative GRMHD

Figure 12. Time-averaged and blurred images at 230 GHz for the cooled model (C3D01RM, top row) and the non-cooled model (NC3RM, bottom row). The synthetic images are averaged over 8000 \( t_g \)–9500 \( t_g \) and blurred by the convolution with a 2D Gaussian filter, at which the full width at half-maximum (FWHM) is 20 \( \mu \text{as} \) (white circle in bottom right of each panel). The size of the images extends to \( \pm 65 \mu \text{as} \), which corresponds to \( \sim 15 r_g \). The columns represent the result with different inclination angles: 50° (left), 70° (centre), and 90° (edge-on, right).

5 CONCLUSIONS

By means of GRMHD simulations and GRRT post-processing, we study the effects of radiative cooling on the dynamics of accretion flows and their resulting spectra. It is generally assumed that radiative cooling is negligible for RIAF discs, which occur at low accretion rates (i.e. \( \dot{M} \ll 10^{-7} \dot{M}_{\text{Edd}} \)). However, the importance of radiative cooling increases with the increasing accretion rate. It is poorly understood what the critical value of the BH mass accretion rate beyond which radiative cooling becomes effective actually is, particularly in 3D.

For the calculation of radiative cooling, we adopt the approximate solution for the advection-dominated accretion disc, which includes bremsstrahlung, synchrotron, and inverse Compton scattering (Esin et al. 1996). We assume that the temperature ratio between ions and electrons, \( T_i/T_e \), is constant or depends on the plasma beta. However, recent studies with particle-in-cell simulations show that the temperature ratio increases over time due to the weak ion–electron thermal coupling (Zhdankin, Uzdensky & Kunz 2020). While it remains unclear if there are other mechanisms for the effective energy transfer from ions to electrons (e.g. the ion cyclotron instability; Sironi & Narayan 2015), the subject of changing temperature ratio over time is beyond the scope of the current paper. In this work, full 3D GRMHD simulations with radiative cooling are extended from previous 2.5D simulations studied by Dibi et al. (2012) and Drappeau et al. (2013).

In general, radiative cooling enhances the disc mid-plane density as the gas pressure decreases due to energy loss, which is radiated away. The disc with reduced pressure and compressed volume increases the dominance of magnetic fields, which reduce the angular momentum transport outwards via the MRI. As a result, when radiative cooling is on, the disc structure is different from when cooling is neglected: a density peak appears near the central BH, and the distance of the peak from the BH increases with the strength of radiative cooling (i.e. mass accretion rate). This difference is negligible when the accretion rate is small; however, when the accretion rate is larger than \( 10^{-8} M_\odot \text{yr}^{-1} \), it becomes apparent (see Fig. 2). Since this rate lies within the range of mass accretion rates for Sgr A*, we argue that cooling losses can affect the dynamical evolution to an appreciable degree.

The effects of radiative cooling on the spectra are visible even for the low accretion rate of \( \dot{M} < 10^{-8} M_\odot \text{yr}^{-1} \): cooling reduces the peak flux in the sub-mm bumps due to the decreased gas temperature. The decreased seed photon by synchrotron at the sub-mm bumps results in the decrease of the flux in the X-ray bumps for which inverse Compton is responsible. The synthetic images at 230 GHz, which is calculated by GRRT post-processing, show similar crescent shapes between the cooled and non-cooled GRMHD data but slightly dimmer in the cooled data due to the decreased temperature adjacent to the BH.

Recent studies by Ressler et al. (2020) indicate that the inner regions near the BH could be strongly magnetized, as magnetic fields get advected from stellar winds (see also Ressler, Quataert & Stone 2018), which can lead to the formation of MADs. While it is thought to be inevitable for MADs to produce strong outflows, which are absent in Sgr A* (e.g. Markoff et al. 2007), we plan to conduct a study of MADs with radiative cooling in future work, so as to investigate how cooling would affect the disc and the outflow in such a situation. Another notable caveat of our simulations is the absence of non-thermal electron acceleration, which is deemed
Figure 13. Light curves at three different frequency bands, from top to bottom: sub-mm \((9.4 \times 10^{11} \text{ Hz})\), near-IR \((4.5 \times 10^{14} \text{ Hz})\), and X-ray \((5 \times 10^{17} \text{ Hz})\), for the cooled model \((\text{C3D1RMh20, solid curve})\) and the non-cooled model \((\text{NC3RM, dashed curve})\). The bottom panel presents the mass accretion rate for the cooled and non-cooled models. The light curves are calculated by using GRMONTY with the assumption of \(D_{\text{Sgr}} = 8.2 \text{ kpc}\), where \(D_{\text{Sgr}}\) is the distance to Sgr A*. The fluxes are averaged over the theta bin of \(60^\circ - 75^\circ\). The bottom panel shows the mass accretion rate at the event horizon. The cadence of the simulation is \(50 \Delta t\), which corresponds to \(\sim 16\) min in Sgr A*. The vertical dotted lines indicate the selected time for a flaring-like event (blue) and a quiescent state (orange).

Figure 14. Comparison of GRRT synthetic images at 230 GHz for the cooled model \((\text{C3D1RMh20})\) between two different states: NIR-quiescent state at \(t = 53\) h and NIR-flaring state at \(t = 37\) h (see Fig. 13). The inclination angle is \(70^\circ\).

Figure 15. Comparison of SEDs between two different states: NIR-quiescent state at \(t = 53\) h and NIR-flaring state at \(t = 37\) h (see Fig. 13). The inclination angle is \(70^\circ\).

Figure 16. Dimensionless mass accretion rate as a function of time for the non-cooled high-resolution 2.5D model \((\text{NC2RH, orange})\) and the non-cooled medium-resolution 3D model \((\text{NC3RM, blue})\).

to be responsible for X-ray (and perhaps, NIR) flaring in Sgr A* (Neilsen et al. 2013; Ball et al. 2016; Connors et al. 2017). We will discuss the contribution of the non-thermal electron in a different work (Chatterjee et al., in preparation).

ACKNOWLEDGEMENTS

This research was enabled in part by support provided by Oak Ridge Leadership Computing Facility, which is a DOE office of Science User Facility supported under contract DE-AC05-00OR22725, and Calcul Québec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca). DY, KC, and SM are supported by the Netherlands Organization for Scientific Research (NWO) VICI grant (no. 639.043.513), ZY is supported by a Leverhulme Trust Early Career Research Fellowship, ML was supported by the NWO Spinoza Prize (PI M.B.M. van der Klis), and AT is supported by Northwestern University and by National Science Foundation grants AST-1815304 and AST-1911080.
APPENDIX: VALIDATION OF SPECTRAL CALCULATION WITH GRMONTY

We make use of BHOSS to reproduce the synthetic images at 230 GHz, while GRMONTY is used to calculate the broad-band spectra since the radiative processes in BHOSS include only the synchrotron emission and absorption. To verify consistency between the two codes, we compare the sub-mm bump in the calculated spectra, as the synchrotron emission is dominant in the sub-mm bump. Fig. A1 shows good agreement for the frequency range of $10^{11}$ Hz $< \nu < 2 \times 10^{13}$ Hz, within which the bump is located.

GRMONTY is known to converge to the correct solution as the fractional error $\propto N_{sp}^{-1/2}$ for the optically thin synchrotron sphere, where $N_{sp}$ is the number of the super-photons (Dolence et al. 2009). Evidently, the spectra with different $N_{sp}$ are consistent with each other while the spectra calculated with smaller $N_{sp}$ exhibit more fluctuations at high frequencies than spectra calculated with larger $N_{sp}$. Due to limited computing resources, we choose the number of super-photons as $N_{sp} = 5 \times 10^5$ for the series of snapshots ($\sim 100$ snapshots for a single run), and to reduce the sampling fluctuations due to small $N_{sp}$, we smooth the spectra using a 1D Gaussian filter with $\sigma = 1$. Given that the number of data points is 200, the size of the energy bin is $1.5 \times 10^{33}$ erg s$^{-1}$ and the value of $\sigma = 1$ corresponds to the FWHM of $3.5 \times 10^{33}$ erg s$^{-1}$. Fig. A1 shows the difference of the resulting spectra with the different $N_{sp}$. In bottom panel, we calculate the fractional difference, which is expressed as,

$$\frac{\delta(v \mathcal{L}_v)}{v \mathcal{L}_v} = \frac{L_{v,Nsp5} - L_{v,Nsp5,smoothed}}{L_{v,Nsp5}},$$

where $L_{v,Nsp5,smoothed}$ and $L_{v,Nsp5}$ are the luminosities, which are calculated with $N_{sp} = 5 \times 10^5$ super-photons and a smoothing process, and exclusively with $10^5$ super-photons, respectively. As seen in the figure, the fractional difference is small ($\delta(v \mathcal{L}_v)/v \mathcal{L}_v << 1$) for all frequency ranges, and thus we are confident in calculating spectra with $N_{sp} = 5 \times 10^5$. However, this may not be sufficient for the optically thick synchrotron sphere since the correct computation of the photon-weights in the large optical depth regime requires a minimum number of super-photons for convergence (Dolence et al. 2009). The current public release of GRMONTY is available only with Open-MP, which works with multiple processors in a single node. In future work, especially studies investigating the case of high accretion rates, it is necessary to incorporate acceleration schemes such as MPI-parallelization to be able to use a large number of super-photons.

Figure A1. Comparison of spectra calculated in BHOSS (blue curve) and GRMONTY. Upper panel: orange and red curves represent the spectra calculated by GRMONTY with $N_{sp} = 5 \times 10^5$ and $10^7$ super-photons, respectively. The green curve represents the smoothed spectrum with $N_{sp} = 5 \times 10^5$ super-photons, where a Gaussian filter with $\sigma = 1$ is employed. Bottom panel: fractional difference of spectra between spectrum with a large number of super-photons (red curve) and a smaller number super-photons with the aforementioned smoothing process (green curve), calculated using equation (A1).