Chapter 5
Post-Merger Jets from SMBH Coalescences as Electromagnetic Counterparts of GW Emission

Note: The material in this Chapter is based on my paper [363], with co-authors Kohta Murase, B. Theodore Zhang, Shigeo S. Kimura, and Peter Mészáros.

5.1 Introduction

Supermassive black hole (SMBH) mergers are ubiquitous in the history of the Universe [75, 76,364] and can produce powerful gravitational wave (GW) bursts when they coalesce [365, 366], making them promising candidates for GW detectors such as Laser Interferometer Space Antenna (LISA, [77,367]) and pulsar timing arrays (PAT, [368–370]) in single-source and/or stochastic GW background searches. The accretion activity between the binary system and the surrounding disk can produce multi-wavelength electromagnetic (EM) emission [79, 320, 371–374], and the time-variable EM signatures from the circumbinary disks could be detectable [375–377]. The spinning SMBH expected to form after the SMBHs have coalesced may also lead to relativistic jets, in which particle acceleration will take place. The resulting non-thermal emission from the accelerated electrons may provide a promising post-merger EM counterpart of the GW emission, and will not only provide complementary information on SMBH mergers but also shed light on the physical processes in these systems [378–380]. [381] recently suggested that the SMBH mergers can also be high-energy neutrino emitters, and demonstrated that they are also promising targets for high-energy multi-messenger astrophysics [274].

We study the EM emission produced in relativistic jets launched after the coalescence

of SMBHs. The physical picture is that the disk winds originating from the circumbinary disk and mini-disks around each SMBH form a pre-merger wind bubble, and jets powered by the Blandford-Znajek (BZ, [80]) mechanism are launched after the merger. The jets push ahead inside the pre-merger disk wind material, resulting in the formation of forward and reverse shocks. In the forward shock region, electrons are accelerated to high energies with a power-law distribution as observed in afterglows of gamma-ray bursts (GRBs) [81]. These particles then produce broadband non-thermal EM emission through synchrotron and synchrotron self-Compton (SSC) processes.

This letter is organized as follows. In Sec. 5.2, we introduce the physical conditions of the pre-merger wind bubble and model the propagation of jets. The radiation processes and the resulting photon spectra, light curves and detection horizons are presented in Sec. 5.3. In Sec. 5.4, we discuss implications of our results. Throughout the letter, we use the conventional notation $Q_x = Q/10^x$ and physical quantities are written in the centimeter-gram-second units, unless otherwise specified.

5.2 Jet Dynamics

We discuss here the physical conditions in a pre-merger circumbinary environment and derive relevant quantities that describe the jet propagation. We consider on-axis observers, which is sufficient for the purpose of this work. The emission region is typically expected to be only mildly relativistic on time scales of interest (the corresponding observation time after the jet launch is $T \sim 10^5 - 10^6$ s).

Numerical simulations have demonstrated that binary SMBH mergers can produce jet-like emissions driven by the Poynting outflow [372]. We assume that a jet is launched after the coalescence and subsequently propagates in the wind bubble formed by pre-merger disk winds. Fig. 5.1 schematically illustrates the configuration of the system. The disk wind expands in the gaseous environment of the host galaxy. We focus on emissions from the shock between the jet and the wind bubble. Initially, the circumbinary disk can react promptly to the evolution of the binary system. The ratio between the disk radius R_d and the semi-major axis of the binary system a remains unchanged ($R_d/a \sim 2$), until the inspiral time scale $t_{\rm GW}$ of the binary system [319] equals the viscosity time scale $t_{\rm vis}$ [317], which is known as the disk decoupling. After the disk becomes decoupled, the merger of SMBHs in binary system occurs within the time interval $t_m \sim 3 \times 10^{-2}$ yr $M_{\rm BH,6}\alpha_{-1}^{-8/5}h_{-1}^{-16/5}$, where $M_{\rm BH} = 10^6\ M_{\rm BH,6}M_\odot$ is the mass of the binary system, the dimensionless parameter h is defined as $h = H/R_d$, $\alpha \sim 0.1$ is the

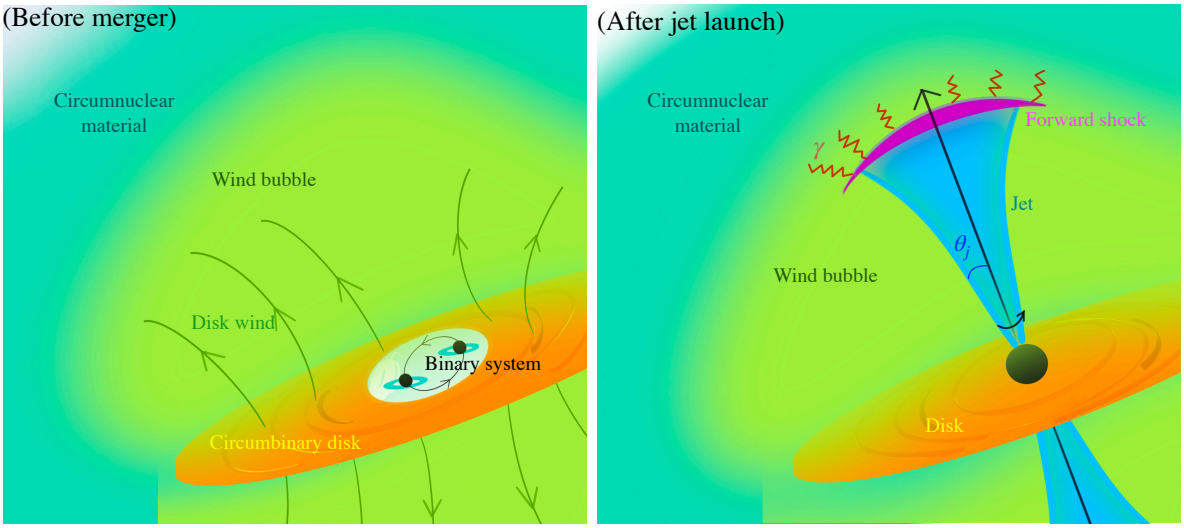

Figure 5.1. Schematic description of our model. **Left panel:** pre-merger disk winds launched from the circumbinary disk. The green arrows illustrate the disk-driven outflows that form a wind bubble. Mini-disks around each SMBH are also shown. **Right panel:** post-merger jets launched by a merged SMBH. The forward shock region is shown as the purple area. The cocoon is not depicted.

viscosity parameter, and H is the disk scale height. The disk gas starts to fill the cavity between the disk and the SMBHs in the viscosity timescale,

$$t_{\rm vis} \sim 0.1 \text{ yr } M_{\rm BH,6}\alpha_{-1}^{-8/5}h_{-1}^{-16/5} \tag{5.1}$$

after the coalescence [320]. This leads to a time delay ($t_{\rm delay} \sim t_{\rm vis}$) of days to months between the GW burst and the launch of post-merger jets, if $h \sim 0.1 - 0.3$ is assumed. However, for a thick and highly magnetized disk with $h \sim \alpha \sim 1$, $t_{\rm delay}$ could be much shorter.

On the other hand, within the duration of these two short-term processes, e.g., $t_{\rm m}$ and $t_{\rm vis}$, the disk wind radius may reach $v_d(t_{\rm vis} - t_m) \sim 10^{14} - 10^{16}$ cm above the disk, where v_d is the disk wind velocity that is of the order of the escape velocity, $v_{\rm esc}(R_{d,\rm dec}) \approx \sqrt{2GM_{\rm BH}/R_{d,\rm dec}}$ for the circumbinary disk, and $R_{d,\rm dec} \approx 1.2\times10^{13}$ cm $M_{\rm BH,6}\alpha_{-1}^{-2/5}h_{-1}^{-4/5}$ is the radius of the circumbinary disk at the decoupling. In reality, not only the circumbinary disk but also mini-disks around two SMBHs contribute, which would make the wind bubble more complicated. For simplicity, we assume the density profile of the winds at the decoupling to obtain the density distribution of the wind bubble at larger distances,

$$\varrho_w(r) = \frac{\eta_w(1+\chi)\dot{M}_{\rm BH}}{4\pi r^2 v_d} \equiv Dr^{-2}, \tag{5.2}$$

where $\dot{M}_{\rm BH}$ is the mass accretion rate onto the binary system, $\chi \sim 1$ is introduced to take into account the contribution of mini-disks, and η_w represents the fraction of accreted mass converted to the disk wind. According to the simulations, for SANE (Standard And Normal Evolution) models, the parameter η_w may vary from 10^{-4} to 10^{-1} [382–384] when the mass accretion rate changes from sub-Eddington to super-Eddington. In MAD (Magnetically Arrested Disk) models, η_w can reach 10^{-2} to 10^{-1} [385]. With $v_d \sim v_{\rm esc}(R_{d,\rm dec})$, we have $D \simeq 5.9 \times 10^{11}$ g cm^{-1} $\tilde{\eta}_{w,-1.5}(\dot{m}/0.5)M_{\rm BH,6}\beta_{d,-1}^{-1}$, where $\tilde{\eta}_w \equiv (1+\chi)\eta_w$, $\beta_{d,-1} \equiv v_d/(0.1c)$, the parameter $\dot{m}$ is defined as the ratio of $\dot{M}_{\rm BH}$, and the Eddington value $\dot{M}_{\rm Edd} \equiv 10L_{\rm Edd}/c^2$ (assuming a radiation efficiency of 0.1).

After the coalescence, a powerful jet driven by the spin energy of the newly formed SMBH can appear, subsequently propagating in the pre-merger wind bubble. Considering a sub-Eddington accretion rate with the MAD configuration, we estimate the jet kinetic luminosity to be

$$\begin{aligned} L_{k,j} &= \eta_j \dot{M}_{\rm BH} c^2 \\ &\simeq 6.3 \times 10^{44} \text{ erg s}^{-1} \eta_j (\dot{m}/0.5) M_{\rm BH,6}, \end{aligned} \tag{5.3}$$

where $\eta_j \sim 0.3 - 1$ is the ratio of the accretion energy converted to the jet energy [325].

Following the standard jet propagation theory [306, 307], we write down the dimensionless parameter that represents the ratio of the energy density of the jet and the rest-mass energy density of the surrounding medium

$$\tilde{L} \approx \frac{L_{k,\rm iso}}{4\pi r^2 \varrho_w c^3} \simeq 63\ \tilde{\eta}_{w,-1.5}^{-1} \eta_j \theta_{j,-0.5}^{-2} \beta_{d,-1}, \tag{5.4}$$

where θ_j is the jet opening angle, $L_{k,\rm iso} \approx 2L_{k,j}/\theta_j^2$ is the isotropic-equivalent luminosity. Since the quantity $\tilde{L}$ lies in the regime $\theta_j^{-4/3} \ll \tilde{L}$, we expect that the jet is "uncollimated" for our fiducial parameters[1]. This situation is similar to that in choked jet propagation in the circumstellar material [386, 387], and β_h is evaluated from $\tilde{L}$. In the relativistic limit, the jet head Lorentz factor is $\Gamma_h \approx \tilde{L}^{1/4}/\sqrt{2}$ [386], and we have $\Gamma_h \sim 2$ in our fiducial case with $\tilde{\eta}_w = 10^{-1.5}$. Note that the jet head radius is $R_h = c\beta_h \hat{T} \approx c\hat{T}$, and $\hat{T}$ is introduced to represent time measured in the central engine frame, which can be converted to the observation time T via $T = (1+z)(1-\beta_h)\hat{T}$ (that is $T \approx (1+z)\hat{T}/[2\Gamma_h^2]$ in the relativistic limit) for on-axis observers.

Furthermore, to ensure particle acceleration, we impose radiation constraints requiring

[1]However, jet collimation, which was assumed in [381], would be achievable for the super-Eddington accretion accompanied by disk winds with $\eta_w \sim 0.1 - 0.3$.

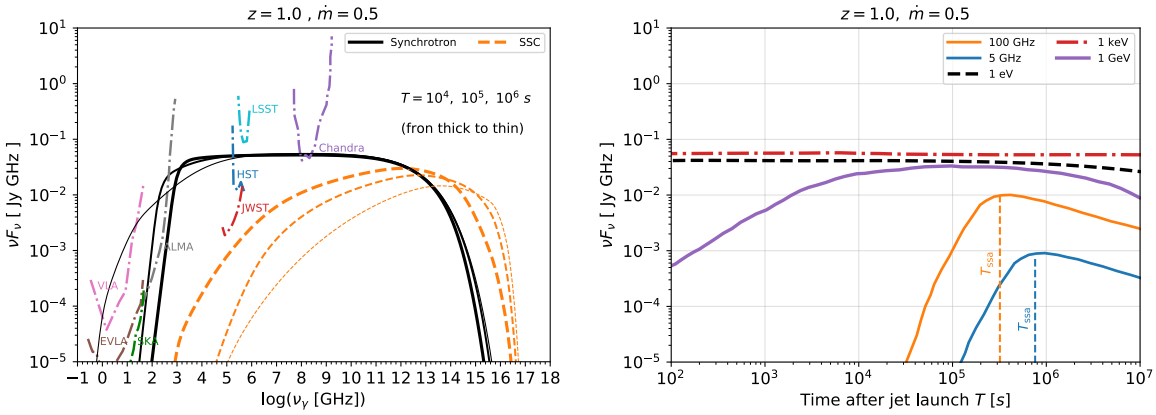

Figure 5.2. Left panel: Non-thermal energy spectra expected for uncollimated post-merger jets from a SMBH merger located at $z = 1$. The solid and dashed lines represent the synchrotron and SSC components. The dash-dotted lines show the sensitivity curves for current and future detectors. **Right panel:** Multi-wavelength light curves. The yellow and blue dashed vertical lines illustrate respectively the characteristic times, e.g., $T_{\rm ssa}$, of 100 GHz and 5 GHz emissions. The used parameters are $\dot{m} = 0.5$, $M_{\rm BH} = 10^6\ M_\odot$, $\tilde{\eta}_w = 10^{-1.5}$, $\eta_j = 1$, $\theta_j = 10^{-0.5}$, $s = 2.0$, $\zeta_e = 0.4$, $\epsilon_e = 0.1$ and $\epsilon_B = 0.01$.

that the shock is collisionless, without being mediated by radiation [386, 388]. Here, ignoring effects of pair production, we use the conservative condition, $\tau_T \approx \varrho_w \sigma_T R_h / m_p < 1$, where σ_T is the Thomson cross section. Numerically, this condition is satisfied at $\hat{T} \gtrsim 10$ s, which is much shorter than the duration of EM emission.

5.3 Electromagnetic Emission from Post-Merger jets

With the jet dynamics presented in the previous section, we calculate the EM spectra resulting from synchrotron and SSC emission. As in the standard theory of GRB afterglows [81], we assume that electrons are accelerated at the external forward shock with a power-law spectral index s. The energy fractions of the downstream energy density converted to non-thermal electron and magnetic field energy are defined as ϵ_e and ϵ_B, respectively. The upstream number density is given by $n_{h,u} \approx \varrho_w(R_h)/m_p \propto R_h^{-2}$, and $B \approx [\epsilon_B 32\pi\Gamma_h(\Gamma_h - 1)n_{h,u}m_p c^2]^{1/2}$ is the downstream magnetic field strength.

In the relativistic limit ($\Gamma_h \gg 1$), the characteristic injection frequency ν_m and the

cooling frequency ν_c in the observer frame are written respectively as,

$$\begin{aligned}\nu_m &\approx \frac{3\Gamma_h\gamma_m^2 eB}{4\pi(1+z)m_e c} \\ &\simeq 3.4\times10^3\ \mathrm{GHz}\ \epsilon_{e,-1}^2\zeta_{e,-0.4}^2\epsilon_{B,-2}^{1/2}\eta_j^{1/2}T_4^{-1} \\ &\quad\times(\dot{m}/0.5)^{1/2}M_{\rm BH,6}^{1/2}\theta_{j,-0.5}^{-1}\end{aligned} \tag{5.5}$$

and

$$\begin{aligned}\nu_c &\approx \frac{3\Gamma_h\gamma_c^2 eB}{4\pi(1+z)m_e c} \\ &\simeq 4.6\times10^2\ \mathrm{GHz}\ (1+z)^{-2}(1+Y)^{-2}\epsilon_{B,-2}^{-3/2} \\ &\quad\times\tilde{\eta}_{w,-1.5}^{-2}\theta_{j,-0.5}^{-1}(\dot{m}/0.5)^{-3/2}T_4\beta_{d,-1}^2 M_{\rm BH,6}^{-3/2},\end{aligned} \tag{5.6}$$

where $\gamma_m = \epsilon_e\zeta_e(\Gamma_h-1)m_p/m_e$ is the electron minimum Lorentz factor, and $\gamma_c = 6\pi m_e c/[(1+Y)T'\sigma_T B^2]$ is the cooling Lorentz factor. Here, $\zeta_e = g_s/f_e = 1/[f_e\ln(\gamma_M/\gamma_m)]\sim 0.3-0.4$ is constrained by the particle-in-cell simulations [213] (where f_e is the fraction of accelerated electrons and the maximum Lorentz factor of electrons is $\gamma_M = (6\pi e)^{1/2}/[\sigma_T B(1+Y)]^{1/2}$), Y is the Compton parameter, and $T' = \hat{T}/\Gamma_h \approx 2\Gamma_h T/(1+z)$ is the comoving time. For example, at $T = 10^4$ s, we have $Y\simeq 2.4$, corresponding to the fast cooling regime. It changes to the slow cooling regime on a time scale from days to weeks. We obtain the peak synchrotron flux [389]

$$\begin{aligned}F_{\nu,\rm syn}^{\rm max} &\approx \frac{(1+z)(0.6 f_e n_{h,u}R_h^3)\Gamma_h e^3 B}{\sqrt{3}m_e c^2 d_L^2} \\ &\simeq 0.24\ \mathrm{mJy}\ (1+z)g_{s,-1.2}\zeta_{e,-0.4}^{-1}(\dot{m}/0.5)^{3/2} \\ &\quad\times\eta_j^{1/2}\epsilon_{B,-2}^{1/2}\tilde{\eta}_{w,-1.5}\beta_{d,-1}^{-1}\theta_{j,-0.5}^{-1}M_{\rm BH,6}^{3/2}d_{L,28}^{-2}.\end{aligned} \tag{5.7}$$

The low-frequency synchrotron emission is subject to synchrotron self-absorption (SSA). The SSA optical depth is written as $\tau_{\rm ssa}(\nu) = \xi_s e n_{h,u}R_h(\nu/\nu_n)^{-p}/[B\gamma_n^5]$, where ν is the observed frequency, $\xi_s\sim 5-10$ depends on the electron spectral index, $\gamma_n = \min[\gamma_m,\ \gamma_c]$, $\nu_n = \gamma_n^2 eB/[(1+z)m_e c]$, $p = 5/3$ for $\nu<\nu_n$ and $p = (4+s)/2$ or $p=3$ for $\nu>\nu_n$ depending on the slow or fast cooling regime (e.g., [331, 390]). The critical time scales set by $\tau_{\rm ssa}=1$ for $\nu<\nu_n$ and for $\nu>\nu_n$ are $T_{\rm ssa}\simeq 5.4\times10^5$ s $\xi_{s,1}^{3/10}(1+z)^{1/2}(1+Y)^{1/2}\epsilon_{B,-2}^{3/5}\left(\frac{\nu}{1\ \rm GHz}\right)^{-1/2}\tilde{\eta}_{w,-1.5}^{11/10}\beta_{d,-1}^{-11/10}(\dot{m}/0.5)^{9/10}M_{\rm BH,6}^{9/10}$ and $T_{\rm ssa}\simeq 3.5\times10^5$ s $\xi_{s,1}^{1/2}(1+z)^{-1/2}(1+Y)^{-1/2}M_{\rm BH,6}^{1/2}\left(\frac{\nu}{100\ \rm GHz}\right)^{-3/2}(\dot{m}/0.5)^{1/2}M_{\rm BH,6}^{1/2}\tilde{\eta}_{w,-1.5}^{1/2}\beta_{d,-1}^{-1/2}$, respectively. Thus, we expect that EM emission at 5 GHz and 100 GHz reaches a peak about a few days

after the jet launch ($T_{\rm ssa} \simeq 7.1 \times 10^5$ s and $T_{\rm ssa} \simeq 3.1 \times 10^5$ s, respectively, in our fiducial case with $\xi_s = 8.7$).

We numerically calculate the electron distribution and the resulting synchrotron and SSC spectra of the forward shock, following the method used in [234] and [391]. We solve the continuity equation that describes the evolution of the electron spectra and calculate the synchrotron/SSC components, in which the trans-relativistic regime can be consistently treated as in [391]. Combining the obtained radio, millimeter, optical and X-ray light curves with the sensitivities of corresponding detectors, we discuss the possibility of follow-up observations of the EM counterpart.

The left panel of Fig. 5.2 shows the snapshots of synchrotron and SSC spectra at $T = 10^4 - 10^6$ s for an on-axis source located at $z = 1$. We assume $s = 2.0$, $\epsilon_e = 0.1$, and $\epsilon_B = 0.01$. The solid and dashed lines correspond to the synchrotron and SSC components. Very high-energy gamma-ray emission at $\gtrsim 1$ TeV energies is suppressed due to the Klein-Nishina effect [234, 391], and the $\gamma\gamma$ annihilation with the extragalactic background light (EBL). For the EBL correction, $\gamma\gamma$ optical depth data from Model C in [185] is used. To show how the EM signal evolves with time, we illustrate the gamma-ray (1 GeV), X-ray (1 keV), UV (1 eV) and radio (5 GHz and 100 GHz) light curves in the right panel. In particular, before the characteristic time $T_{\rm ssa}$ (shown as the vertical yellow and blue lines). The radio emission is suppressed by the SSA process, which is implemented by multiplying $(1 - e^{-\tau_{\rm ssa}})/\tau_{\rm ssa}$.

It is useful to discuss the detection horizon $d_{\rm lim}$ for some detectors such as the Square Kilometre Array (SKA), Very Large Array (VLA), Expanded VLA (EVLA), Atacama Large Millimeter Array (ALMA), Hubble Space Telescope (HST), James Webb Space Telescope (JWST), Large Synoptic Survey Telescope (LSST) and the high-resolution camera on the Chandra X-ray Observatory (Chandra) [2] as functions of the observation time T. Given the observed flux $F_\nu(\nu_\gamma, T, z)$ at the observer time T from an on-axis source located at redshift z, the horizon can be calculated iteratively via

$$d_{\rm lim}(\nu_\gamma, T) = d_L \left(\frac{\frac{1}{\Delta T_{\rm exp}} \int_T^{T+\Delta T_{\rm exp}} F_\nu(\nu_\gamma, t, z) dt}{F_{\rm lim}(\nu_\gamma, \Delta T_{\rm exp})} \right)^{1/2} , \tag{5.8}$$

where $F_{\rm lim}(\nu_\gamma, \Delta T_{\rm exp})$ is the detector sensitivity normalized to the exposure time $\Delta T_{\rm exp}$.

[2] For information on these facilities see, e.g., : VLA (http://www.vla.nrao.edu), EVLA (http://www.aoc.nrao.edu/evla/), SKA (https://www.skatelescope.org), ALMA (https://public.nrao.edu/telescopes/alma/), HST (https://www.nasa.gov/mission_pages/hubble/main/index.html), JWST (https://stsci.edu/jwst), LSST (https://www.lsst.org/scientists/scibook) and Chandra (https://cxc.cfa.harvard.edu/cdo/about_chandra/)

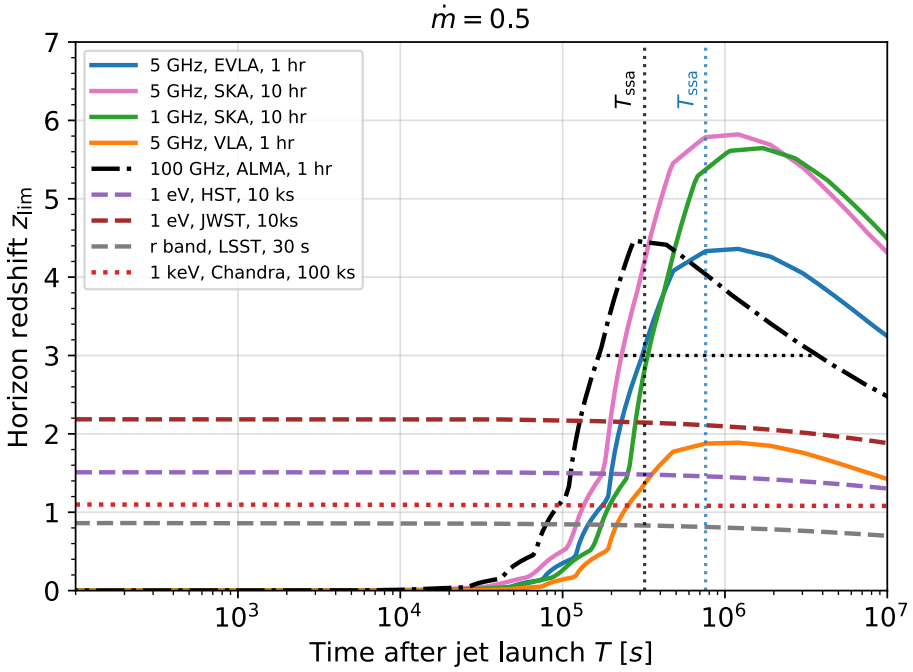

Figure 5.3. Detection horizons for multi-wavelength detectors, e.g., SKA, VLA, EVLA, ALMA, HST, JWST, LSST and Chandra. The horizontal dotted line shows the 100 GHz detection window for ALMA assuming a source located at $z = 3$. Similar to Fig. 5.2, the dotted vertical lines are the characteristic times of 5 GHz and 100 GHz signals.

For example, specifying the detection frequency $\nu = 100$ GHz, the sensitivity of ALMA is approximately 34 μJy for one-hour integration, e.g., $\Delta T_{\rm exp} = 1$ hour. Fig. 5.3 indicates the detection horizons for SKA (5 GHz, $\Delta T_{\rm exp} = 10$ hr), SKA (1 GHz[3], $\Delta T_{\rm exp} = 10$ hr), VLA (5 GHz, $\Delta T_{\rm exp} = 1$ hr), ALMA (100 GHz, $\Delta T_{\rm exp} = 1$ hr), JWST (1 eV, $\Delta T_{\rm exp} = 10$ ks), HST (1 eV, $\Delta T_{\rm exp} = 10$ ks), LSST (r-band, point source exposure time $\Delta T_{\rm exp} = 30$ s in the 3-day revisit time), and Chandra (1 keV, $\Delta T_{\rm exp} = 100$ ks). The vertical black and blue dotted lines respectively illustrate the times $T_{\rm ssa}$ at which photons at 100 GHz and 5 GHz bands start to survive from the synchrotron self-absorption.

From Fig. 5.3, we expect that ALMA, SKA and EVLA can detect SMBH mergers in the radio bands respectively out to redshifts of $z \sim 4 - 6$. Remarkably, the optical and X-ray signals from the mergers in the range $1 \lesssim z \lesssim 2$ can also be identified through targeted searches by Chandra, HST and JWST in a long duration after the merger. In addition, we can estimate the observation time for each detector if the luminosity distance of the merger is specified. For example, a source located at $z = 3$ would remain

[3]At 1 GHz, the SKA field-of-view can reach $\gtrsim$1 deg^2

detectable by ALMA for roughly 20-30 days (see the black dotted horizontal line). One caveat is that this calculation is carried out in the ideal case where the detectors can point to the position of the source and start the observation immediately after the EM signal reaches the Earth. We discuss the sky coverage and a detection strategy in the following Sec. 5.4.

5.4 Summary and Discussion

We investigated broadband non-thermal EM emission from electrons accelerated at the external forward shock expected in post-merger jets from the coalescence of SMBHs. In our model, the jets can be launched at $t_{\rm delay} \sim t_{\rm vis} \sim (0.003 - 0.1) M_{\rm BH,6}$ yr after the coalescence. The time lag is primarily determined by the scale height of the circumbinary disk and the viscosity parameter. We found that, for a moderate accretion rate ($\dot{m} \sim 0.5$), the multi-wavelength emission from such a system may persist at detectable levels for months after the jet launch, depending on the facilities and the luminosity distance. Moreover, according to our model, the sources with moderate $\dot{m} = 0.5$ can be detected up to $z \sim 5 - 6$, covering the range that LISA-like GW detectors have the best detection chance, e.g., $z \sim 1{-}2$, in which $(1{-}10) f_b$ mergers per year are expected [340,345,392–395]. Here $f_b \sim 1/(2\Gamma_h^2)$ is the beaming factor in our model. Because the jet head Lorentz factor is as low as $\Gamma_h \lesssim 2$, the EM emission from the forward shock region is not highly beamed and we expect $f_b \sim 0.1 - 1$. This makes the binary SMBH mergers interesting targets for future multi-messenger studies. If super-Eddington accretion (e.g., $\dot{m} \sim 10$) occurs, as was optimistically assumed in [381], even LSST and Chandra could detect EM signals from the sources in the redshift range $4 \lesssim z \lesssim 6$. We showed the case of $s = 2.0$ for the demonstration. If a larger spectral index, e.g., $s \sim 2.2 - 2.4$, is used, as expected from observations of GRB afterglows, the radio detection would be more promising whereas a higher accretion rate would be required for successful optical and X-ray observations.

The density of the premerger bubble, which was assumed to be a wind profile, is subject to large uncertainties. The extrapolation in the density distribution would be applicable up to an outer wind radius of $\sim 10^{14} - 10^{16}$ cm. The density predicted by equation 5.2 would drop below that of the central molecular zone (indicated as the circumnuclear environment in Fig.5.1), which may lead to the increase of radio emission. In addition, a cocoon formed along with the jet, depending on uncertain details of the medium, could produce thermal photons which may not only lead to detectable signals

but also serve as seed photons for inverse-Compton emission. We focused on the more secure EM emission from the forward shock region as the jet propagates in the wind. In this sense, our prediction for the fluxes are conservative.

EM emission from the external reverse shock and internal shocks can also be expected [396, 397]. Qualitatively, the ratio between the peak fluxes of the reverse and forward shock emission depends on the value of Γ_j, and the reverse shock contribution might be important for $\Gamma_j \gg \Gamma_h$.

Previous studies based on general relativistic three dimensional magnetohydrodynamics simulations have shown that the circumbinary disk and the corona can emit light in UV/EUV bands [374], while X-ray and infrared emission from the post-merger circumbinary disk are expected to last for years [79, 375]. In the pre-merger phase, the orbits of dual SMBH cores may be identified by radio facilities such as VLBI [398]. Blind searches could identify radio or UV/EUV sources from the binary SMBH systems, which would provide complementary constraints on the source location, the accretion rate and the ambient gaseous environment.

Our model can provide a guidance, including the onset times and the detection windows, in developing detection strategies for future EM follow-up observations, once GW signals are detected. Considering the large uncertainties in the localization with GW detectors, an initial follow-up using large field-of-view (FOV) telescopes, like SKA and LSST, would be necessary to more precisely localize the position of the source. After that, we can use the putative positional information from the initial follow-up imaging to guide the observation of narrower FOV telescopes. In particular, for high-redshift mergers in the range $z \sim 2-5$, EM follow-up observations rely more on radio detectors, and the detection is possible a few weeks after the merger. SKA needs the source localization before follow-up observations by VLA and ALMA. On the other hand, if the merger is close enough (e.g., $z \sim 1$), LISA observations staring from a few weeks before the merger can localize the merger with a median precision of $\sim$1 deg^2 [399]. In this case, LISA and LSST can jointly guide other X-ray and optical facilities in the very early stage. Amid these two regimes, e.g., $z \sim 1-2$, detections in the optical and X-ray bands using HST, JWST and Chandra would be promising if the source is localized by SKA.

Acknowledgements

C.C.Y. and P.M. acknowledge support from the Eberly Foundation. The work of K.M. is supported by NSF Grant No. AST-1908689, and KAKENHI No. 20H01901 and

No. 20H05852. B.T.Z. acknowledges the IGC fellowship. S.S.K. acknowledges the JSPS Research Fellowship, JSPS KAKENHI Grant No. 19J00198.

Disclaimer

The findings and conclusions do not necessarily reflect the view of the funding agencies.

Chapter 6
GeV Signatures of Short GRBs in Active Galactic Nuclei

Note: The material in this Chapter is based on my paper [400], with co-authors Kohta Murase, Dafne Guetta, Pe'er Asaf, Imre Bartos, and Peter Mészáros.

6.1 Introduction

As one of the most luminous and energetic phenomena in the universe, gamma-ray bursts (GRBs) have fueled a vibrant field of astrophysics research for several decades. Based on the duration of the bursts, the population can be divided into two subclasses, long GRBs and short GRBs, which are thought to arise from different progenitors. The general view is that short GRBs result from compact binary object (CBO) mergers [82–87], such as binary neutron star mergers and potentially NS-black hole mergers, whereas long GRBs are generated during the death of massive stars [88–93]. In 2017, the coincident detection of gravitational waves (GWs) and the corresponding electromagnetic counterpart from the binary neutron star merger GW170817, located in the host galaxy NGC 4933, marked a triumph of multi-messenger astronomy [94–96]. The spatial and temporal association between GW170817 and the gamma-ray burst GRB 170817A also consolidates the theory that CBO mergers are the origin of short GRBs. Extensive efforts have shown that the broadband emission is consistent with a relativistic jet viewed from an off-axis angle [96–106]. Moreover, [107] investigated the upscattered cocoon emission as the source of the γ-ray counterpart. The long-lasting high-energy signatures of the central engine left after the coalescence was studied in [108].

Alternatively, unlike in the case of GW170817, one can expect a sub-population of short GRBs which occur in the accretion disks of AGNs. Studies of the CBO formation and

evolution in AGN disks demonstrate that hierarchical mergers of embedded binary black hole systems are promising for reconstructing the parameters of LIGO/VIRGO detected mergers [109–112]. These mergers can harden the black hole mass distribution [113–116] as well. [117] pointed out that mergers involving neutron stars, such as GW190814 and GW190425, could also arise in AGN disks. Recent progress on the optical counterpart to GW190521 could support this [118], although the confirmation needs further observations [119]. [120] systematically studied the electromagnetic signatures of both long GRBs and short GRBs in AGN disks and discussed the conditions for shock breakout. [121] and [122] focused more on the neutrino production of embedded explosions. However, [123] showed that CBO environments are likely to be thin because of outflows that are common in super-Eddington accretion.

In this work, we study γ-ray emission from short GRBs that are embedded in AGN disks. Inside the accretion disk, the embedded objects can migrate towards a migration trap due to angular momentum exchange via the torques originated from the disk density perturbations. At the migration trap, the gas torque changes sign, and an equilibrium is achieved as the outwardly migrating objects meet inwardly migrating objects. Numerical calculations show that compact binaries are typically formed near the migration trap at distances around $R_d \sim 20 - 300 R_{\rm S}$ to the central supermassive black hole (SMBH, [124]), where $R_S = 2GM_\star/c^2$ is the Schwarzschild radius. Employing one-dimensional N-body simulations, [115] obtained a more distant location for typical mergers at $\sim 10^{-2}-10^{-1}$ pc ($\sim 10^3 - 10^4 R_{\rm S}$ for a SMBH with mass $M_\star = 10^8 M_\odot$). We concentrate on the embedded GRBs with distances $R_d \sim 10 - 10^3 R_{\rm S}$. We will show that AGN disks would not influence the γ-ray emission if the CBO mergers happen further outside in the disk. We also note that $R_d = 10 R_{\rm S}$ is an extreme case where the population is stringently limited. The outflows from the binary systems with super-Eddington accretion rates are expected to form a low-density cavity-like structure before the merger occurs [123]. Within such a cavity a successful GRB jet is likely to develop, since the ambient gas density is not sufficiently high to stall the jet, in contrast to the choked-jet case discussed in [122].

In GRB theories, EIC processes can be important when seed photons in the external regions or late/early-time dissipation processes can be efficiently upscattered to the GeV-TeV bands by accelerated electrons [107, 401, 402]. The EIC scenario can be used to explain the observed very-high-energy (VHE) emission from GRBs [403, 404]. In the present case, the disk black body emission provides an appropriate supply of thermal photons to the short GRB jets.

Adopting a thin-disk model, we derive the conditions for cavity formation and

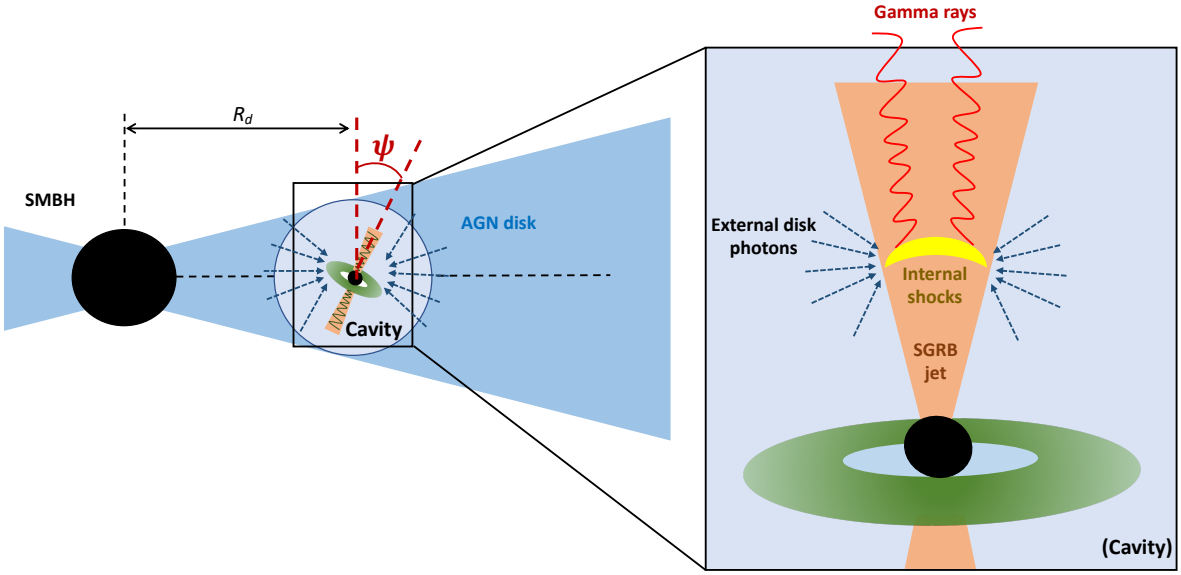

Figure 6.1. Schematic picture of the CBO mergers embedded in AGN disks. A cavity is formed due to the powerful outflows from the circumbinary disk. In this configuration, ψ represents the angle between the CBO orbital plane and the AGN disk, and R_d is the distance between the CBO and the central SMBH. Non-thermal electrons accelerated in the internal dissipation region are responsible for the production of γ-rays. These electrons can upscatter the disk photons, leading to the EIC emission.

calculate disk photon spectra in Sec. 6.2. In Sec. 6.3, we numerically solve the steady-state transport equation to obtain the electron distribution inside the jet. In Sec. 6.4, we calculate the synchrotron, synchrotron self-Compton (SSC) and EIC components. The effects of $\gamma\gamma$ absorption in the AGN disk and electromagnetic cascades are also taken into account. We also present the detection perspectives for the *Fermi* Large Area Telescope (*Fermi*-LAT) and the VHE γ-ray facilities, such as the Major Atmospheric Gamma Imaging Cherenkov (MAGIC), the High Energy Stereoscopic System (H.E.S.S.), the Very Energetic Radiation Imaging Telescope Array System (VERITAS), the Cherenkov Telescope Array (CTA), and the water Cherenkov detector array in the Large High Altitude Air Shower Observatory (LHAASO-WCDA), in Sec. 6.4.2. The prompt emissions are discussed in Sec. 6.4.3. We summarize and discuss the results in Sec. 6.5.

Throughout the paper, we use the notation $Q_x = Q/10^x$, and physical quantities are written in CGS units unless otherwise specified. Quantities with the prime symbol, e.g., Q', are written in the jet comoving frame. We use the symbol $F[a, b, c, ...]$ to represent the value of a function F evaluated at the point $(a, b, c, ...)$.

6.2 Cavity Formation and Disk Photon Spectra

In this section we derive the conditions for the formation of a low-density cavity around the CBO, following the treatment in [123], and model the AGN disk temperature distribution assuming a steady thin disk.

6.2.1 Cavity Formation

For a thin AGN disk with an aspect ratio $h_{\rm AGN} = H_{\rm AGN}/R_d \sim 0.01$ surrounding a SMBH with mass $M_\star = 10^8 M_{\star,8} M_\odot$, we write down the accretion rate onto the SMBH and the radial drift velocity v_R as, respectively, $\dot{M}_\star = \dot{m}_\star L_{\rm Edd,\star}/c^2 \simeq 1.4 \times 10^{25}\ \dot{m}_\star M_{\star,8}$ g s^{-1} and $v_R = \nu/R_d \approx \alpha h_{\rm AGN}^2 v_K \simeq 2.1 \times 10^4\ \alpha_{-1} h_{\rm AGN,-2}^2 \mathcal{R}_2^{-1/2}$ cm s^{-1} [405], where $H_{\rm AGN}$ is the scale height of the AGN disk, $\alpha \sim 0.1$ is the viscous parameter, ν is the kinematic viscosity, $v_K = \sqrt{GM_\star/R_d}$ is the Kepler velocity, R_d is the distance between the CBO and the central SMBH, the dimensionless parameter $\mathcal{R}$ is defined as $\mathcal{R} \equiv R_d/R_{\rm S}$, and $L_{\rm Edd,\star}$ stands for the Eddington luminosity. The surface density for a stable disk can then be written as $\Sigma_{\rm AGN} = \dot{M}_\star/(2\pi R_d v_R) \simeq 3.6 \times 10^4\ \dot{m}_\star M_{\star,8} \mathcal{R}_2^{-1/2} \alpha_{-1}^{-1} h_{\rm AGN,-2}^{-2}$ g cm^{-2}. When a CBO is present in the AGN disk, the surface density is perturbed, and a density gap will appear bracketing the binary's orbit around the SMBH [406]. For a typical short GRB progenitor, we expect the total mass of the binary system to be $M_{\rm CBO} \lesssim 10 M_\odot$. In this case $\Sigma_{\rm CBO} \approx \Sigma_{\rm AGN}$ is a good approximation to the surface density of the AGN disk at the binary's position [123]. We obtain the disk gas density in the vicinity of the CBO

$$\begin{aligned}\rho_{\rm CBO} = \frac{\Sigma_{\rm CBO}}{2H_{\rm AGN}} \simeq & 6.1 \times 10^{-10}\ \dot{m}_\star M_{\star,8} \\ & \times \mathcal{R}_2^{-3/2} \alpha_{-1}^{-1} h_{\rm AGN,-2}^{-3}\ {\rm g\ cm^{-3}},\end{aligned} \tag{6.1}$$

and the disk magnetic field

$$\begin{aligned}B_d &= \sqrt{8\pi\beta^{-1}(\rho_{\rm CBO}/m_p)k_B T_d} \\ &\simeq 2.1 \times 10^2\ \beta_{0.48}^{-1/2} \dot{m}_\star^{1/2} M_{\star,8}^{1/2} R_2^{-3/4} \alpha_{-1}^{-1/2} \\ &\times h_{\rm AGN,-2}^{-3/2} T_{d,5}^{1/2}\ {\rm G},\end{aligned} \tag{6.2}$$

where $\beta \sim 3 - 30$ is define as the ratio of the plasma pressure to the magnetic pressure and T_d is the disk temperature. Henceforth, the sub-index 'CBO' will be used to stand for quantities describing CBOs.

We estimate the accretion rate of the CBO to be $\dot{M}_{\rm CBO} \approx \eta_{\rm CBO} \dot{M}_\star \simeq 1.4 \times$

$10^{24}\ \dot{m}_\star M_{\star,8}\eta_{\rm CBO,-1}$ erg s^{-1}, where $\eta_{\rm CBO}$ is the ratio of the CBO accretion rate to the SMBH accretion rate. This approximation is justified in [123]. We find that the accretion is highly super-Eddington, e.g., $\dot{m}_{\rm CBO} = \dot{M}_{\rm CBO}c^2/L_{\rm Edd,CBO} \simeq 10^6 \dot{m}_\star M_{\star,8} M_{\rm CBO,1}^{-1}\eta_{\rm CBO,-1}$, and expect a wind bubble to be produced by the strong radiation-driven outflows [407–409]. The bubble's expansion in a uniform medium can be described by the formula $r_B \approx 0.88(L_w t^3/\rho_{\rm CBO})^{1/5}$ [410, 411], where r_B is the bubble radius, $L_w = \eta_w \dot{M}_{\rm CBO} v_w^2 \simeq 1.4\times10^{42}\ \dot{m}_\star M_{\star,8}\eta_{\rm CBO,-1}\eta_w v_{w,9}^2$ erg s^{-1} and $v_w \sim 10^9 v_{w,9}$ cm s^{-1} is the outflow velocity. Since the accretion is highly super-Eddington, the factor η_w can reach $\sim 90-100\%$ [412, 413]. However, we use a conservative value $\eta_w \sim 0.3\ \eta_{w,-0.5}$ [414]. Equating the bubble radius r_B to $H_{\rm AGN}/\cos\psi$, we obtain the timescale to create a cavity reaching the approximate boundary of the AGN disk along the direction of the GRB jet,

$$\begin{aligned} t_{\rm cav} &\approx 1.2\left(\frac{\rho_{\rm CBO}H_{\rm AGN}^5}{L_w\cos^5\psi}\right)^{1/3} \\ &\simeq 4.0\times10^5\ (\cos\psi)^{-5}\mathcal{R}_2^{7/6}\alpha_{-1}^{-1/3}h_{\rm AGN,-2}^{2/3} \\ &\quad\times\eta_{\rm CBO,-1}^{-1/3}\eta_{w,-0.5}^{-1/3}v_{w,9}^{-2/3}\ {\rm s}, \end{aligned} \tag{6.3}$$

where ψ is the angle between binary orbital plane and the AGN disk (see the schematic picture in Fig. 6.1). One caveat is that we assumed a spherical outflow to derive the cavity timescale, equation 6.3. [409] pointed out that the outflow is concentrated in a wide-angle funnel that surrounds the jet if the accretion rate is highly super-Eddington. In the following text, we will continue using the spherical cavity timescale for simplicity to obtain sufficient but unnecessary conditions for the cavity formation.

The formation of a cavity for a CBO located at R_d before the merger occurs requires

$$t_{\rm cav} \lesssim \min\left[t_{\rm gw},\ t_{\rm mig},\ t_{\rm vis}\right], \tag{6.4}$$

where $t_{\rm gw}$, $t_{\rm mig}$, $t_{\rm vis}$ are binary merger, migration and AGN disk viscosity timescales,

respectively. We write down the timescales for an equal-mass binary explicitly as

$$\begin{aligned} t_{\rm gw} &= \frac{5}{128}\frac{\dot{m}^4_{\rm CBO}}{A^4_{\rm in}}\frac{GM_{\rm CBO}}{c^3} \\ &\simeq 1.9\times10^{14}\ \dot{m}^4_{\rm CBO,6}A^{-4}_{\rm in,1}M_{\rm CBO,1}\ {\rm s}, \\ t_{\rm mig} &= \frac{h^2_{\rm AGN}M^2_\star}{M_{\rm CBO}R_d v_K \Sigma_{\rm AGN}} \\ &\simeq 1.47\times10^{14}\ \alpha_{-1}h^4_{\rm AGN,-2}M^{1/2}_{\star,8}M^{-1}_{\rm CBO,1}\dot{m}^{-1}_\star\ {\rm s}, \\ t_{\rm vis} &= \frac{R_d}{\alpha h^2_{\rm AGN}v_K} \\ &\simeq 1.39\times10^{11}\ \alpha^{-1}_{-1}h^{-2}_{\rm AGN,-2}\mathcal{R}^{3/2}_2M^{-1/2}_{\star,8}\ {\rm s}, \end{aligned} \tag{6.5}$$

where $A_{\rm in}\sim10$ is the ratio of the inner edge of the circumbinary disk surrounding the CBO and the major axis of the binary's orbit [415]. We define a critical angle ψ_c above which the condition described by equation 6.4 is no longer satisfied, and obtain

$$\psi_c \simeq \frac{\pi}{2} - \max\left[h_{\rm AGN},\ 0.076\mathcal{R}_2^{-1/15}h^{8/15}_{\rm AGN,-2}M^{1/10}_{\star,8}\right]. \tag{6.6}$$

In the equation above, the dependence on the parameters $\alpha,\ \eta_w,\ \eta_{\rm CBO}$ and v_w are not shown, to simplify the notation. Varying $\mathcal{R}$ in the fiducial range $10-10^3$, we estimate the critical angle $\psi_c\simeq85.6^\circ$ and find that ψ_c depends very weakly on $\mathcal{R}$ and $M_\star$. This result supports the argument that in most cases a cavity surrounding the CBO is unavoidable and the jet is not choked, except if the binary orbital plane is perpendicular to the AGN disk [123].

6.2.2 Disk Photon Spectra

The accretion disk can become optically thick to ultraviolet/infrared photons as the plasma gets ionized. We estimate the vertical optical depth, for a fully ionized disk with temperature $T_d\gtrsim10^4$ K,

$$\begin{aligned} \tau_d &\approx \Sigma_{\rm AGN}\kappa_{\rm R} \\ &\simeq 7.2\times10^3\ (1+X)\dot{m}_\star M_{\star,8}\mathcal{R}_2^{-1/2}\alpha^{-1}_{-1}h^{-2}_{\rm AGN,-2}, \end{aligned} \tag{6.7}$$

where $\kappa_{\rm R}\approx0.2(1+X)$ is the Rosseland mean opacity for Thomson scattering and X is the hydrogen mass fraction. Since the disk remains optically thick ($\tau_d>1$) in the range $\mathcal{R}\sim10-10^3$, we use a black-body spectrum to approximate the local photon density

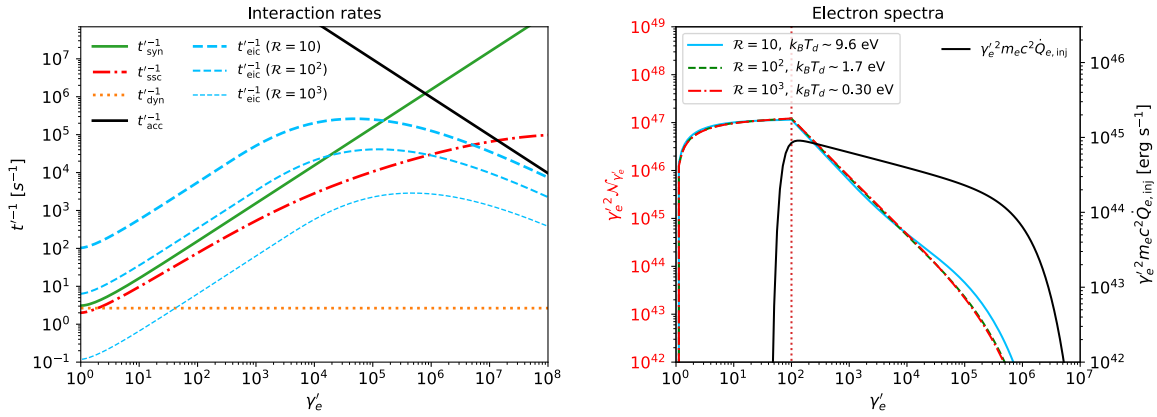

Figure 6.2. *Left panel:* Energy loss rates of accelerated electrons in the internal dissipation region. The green solid and red dash-dotted lines respectively show the synchrotron and SSC rates. From thick to thin, the blue dashed lines depict the EIC cooling rate for the CBOs at $\mathcal{R}=10, 10^2$ and 10^3, respectively. The reciprocals of the dynamic and acceleration times are illustrated as the yellow dotted and black solid lines. *Right panel:* The electron number spectra as functions of the electron Lorentz factor. The minimum injected Lorentz factor is $\gamma'_{e,m}=100$. The blue solid, green dashed and red dash-dotted lines correspond to $\mathcal{R}=10,\ 10^2$ and 10^3 cases. The black solid line is the electron injection function.

(in the units of eV^{-1} cm^{-3}), e.g.,

$$n^{\rm (eic)}_{\varepsilon_\gamma} = \frac{8\pi}{(hc)^3}\frac{\varepsilon_\gamma^2}{\exp\left(\frac{\varepsilon_\gamma}{k_BT_d}\right)-1}, \tag{6.8}$$

where ε_γ is the energy of seed disk photons in the engine frame. The disk temperature T_d at the position of the CBO can be written as [405]

$$\begin{aligned} T_d &= \left\{\frac{2GM_\star\dot{M}_\star}{8\pi\sigma_S R_d^3}\left[1-\left(\frac{R_*}{R_d}\right)^{1/2}\right]\right\}^{1/4} \\ &\simeq 2.0\times10^4\ \dot{m}_\star^{1/4} M_{\star,8}^{-1/4}\mathcal{R}_2^{-3/4}\ {\rm K}, \end{aligned} \tag{6.9}$$

where σ_S is the Stefan-Boltzmann constants and R_* is the innermost edge of the disk. In this paper, we consider three distances $\mathcal{R}=10,\ 10^2$ and 10^3. The corresponding disk temperatures are $k_BT_d=9.1$ eV, 1.7 eV and 0.3 eV. For $R_d\gg R_*$, we have $T_d\propto\mathcal{R}^{-3/4}$, implying that the EIC component becomes increasing important when we move the CBO close to the central SMBH.

6.3 Non-Thermal Electrons

We consider a successful (i.e.non-choked) GRB jet whose extended emission has a luminosity $L_{j,\rm iso} = 10^{48.5}$ erg s^{-1}. We focus on the internal dissipation model in which the jet kinetic energy is dissipated at $R_{\rm dis} = 2\Gamma_j^2 c t_{\rm var} \simeq 1.5\times10^{12}\ \Gamma_{j,1.7}^2 t_{\rm var,-2}$ cm via internal shocks [416] or magnetic reconnections [417], where $\Gamma_j = 50\Gamma_{j,1.7}$ is the jet Lorentz factor, $t_{\rm var} = 10^{-2} t_{\rm var,-2}$ s is the variability time of velocity fluctuations. One necessary condition for electron acceleration is that the upstream region should be optically thin for the shock not to be radiation mediated, namely, $\tau_{\rm in} = n'\sigma_T R_{\rm dis}/\Gamma_j \lesssim 1$ [269, 418, 419], where $n' = L_{j,\rm iso}/(4\pi R_{\rm dis}^2 \Gamma_j^2 m_p c^3) \simeq 9.6\times10^{11}\ L_{j,\rm iso,48.5}\Gamma_{j,1.7}^{-6} t_{\rm var}^{-1}$ cm^{-3} is the comoving number density and σ_T is the Thomson cross section. Explicitly, we write down the optical depth as $\tau_{\rm in} \simeq 1.8\times10^{-2}\ L_{j,\rm iso,48.5}\Gamma_{j,1.7}^{-5} t_{\rm var,-2}^{-1}$, which indicates that efficient electron acceleration is plausible.

To get the electron distribution we numerically solve the steady-state transport equation

$$\frac{\mathcal{N}_{\gamma_e'}}{t'_{\rm dyn}} - \frac{\partial}{\partial\gamma_e'}\left(\frac{\gamma_e'}{t'_{e,c}}\mathcal{N}_{\gamma_e'}\right) = \dot{Q}'_{e,\rm inj}, \tag{6.10}$$

where γ_e' is the Lorentz factor, $\mathcal{N}_{\gamma_e'} = dN_e/d\gamma_e'$ is the differential spectrum, $t'_{\rm dyn} = R_{\rm dis}/(\Gamma_j c)$ is the dynamical time that may represent adiabatic losses or escape, $t'_{e,c}$ represents the electron cooling time scale, and the function $\dot{Q}_{e,\rm inj}$ is the electron injection rate from shock acceleration. Specifying a spectral index $s = 2.2$, e.g., $\dot{Q}'_{e,\rm inj} \propto {\gamma_e'}^{-s}$, we normalize the injection function via $\int d\gamma_e'(\gamma_e' m_e c^2 \dot{Q}'_{e,\rm inj}) = \epsilon_e L_{j,\rm iso}/\Gamma_j^2$. The factor ϵ_e, defined as the fraction of jet kinetic energy that is converted to electrons, is assumed to be $\epsilon_e = 0.1$. The minimum Lorentz factor $\gamma'_{e,m}$ for injected electrons is assumed to be $\gamma'_{e,m} = 100$.

In the dissipation region, the magnetic field is $B'_{\rm dis} = [8\pi\epsilon_B(\Gamma_{\rm rel}-1)n' m_p c^2]^{1/2} \simeq 3.8\times10^4\ \epsilon_{B,-2}^{1/2} L_{j,\rm iso,48.5}^{1/2}\Gamma_{j,1.7}^{-3} t_{\rm var}^{-1/2}$ G, where $\Gamma_{\rm rel} \simeq 5$ is the relative Lorentz factor between the fast and slow shells. The ratio of $B'_{\rm dis}$ to the disk magnetic field B_d is $B'_{\rm dis}/(\Gamma_j B_d) \simeq 3.8\,\mathcal{R}_2^{9/8}\beta_{0.48}^{1/2}$. Here, we focus on the $\mathcal{R}$-dependence of the magnetic fields, using the fiducial values for all other parameters. We use the modulated magnetic field $B' = \max[B'_{\rm dis}, \Gamma_j B_d]$ to calculate the electromagnetic emission in the dissipation region.

The accelerated electrons lose energy through synchrotron, SSC, and EIC processes within the corresponding timescales $t'_{e,\rm syn}$, $t'_{e,\rm ssc}$ and $t'_{e,\rm eic}$. The net cooling timescale is given by $t'_{e,c} = ({t'}_{e,\rm syn}^{-1} + {t'}_{e,\rm ssc}^{-1} + {t'}_{e,\rm eic}^{-1})^{-1}$. Electrons with higher γ_e' cool down faster while a longer acceleration time, e.g., $t'_{\rm acc} = \gamma_e' m_e c/(eB')$, is required to reach such a high energy.

We thus expect a cutoff Lorentz factor $\gamma'_{e,\rm cut}$ determined by the equation $t'_{\rm acc} = t'_{e,c}$, above which electrons cannot accumulate energy due to the rapid radiation. Using these arguments, the injection function for a spectral index $s > 2.0$ can be written as,

$$\dot{Q}'_{e,\rm inj} = \frac{(s-2)\epsilon_e L_{j,\rm iso}}{\Gamma_j^2 m_e c^2 \gamma'^2_{e,m}} \left(\frac{\gamma'_e}{\gamma'_{e,m}}\right)^{-s} \exp\left(-\frac{\gamma'_e}{\gamma'_{e,\rm cut}}\right). \tag{6.11}$$

The photons from the synchrotron process play the role of seed photons in EIC scattering. Therefore, we need a trial electron spectrum, e.g. $\mathcal{N}'^{(0)}_{\gamma'_e} \sim (t'^{-1}_{\rm dyn} + t'^{-1}_{e,\rm syn} + t'^{-1}_{e,\rm eic})^{-1}\dot{Q}'_{e,\rm inj}$, to evaluate $t'_{\rm ssc}$, and solve the differential equation 6.10 iteratively to obtain a convergent solution as in [403].

The left panel in Fig. 6.2 shows the energy loss rates. The blue dashed lines show the EIC cooling rate for $\mathcal{R} = 10,\ 10^2$ and 10^3. The synchrotron (green line) and SSC (red dash-dotted line) cooling rates are not sensitive to the CBO's position, whereas the EIC rate increases as the distance between the CBO and the SMBH reduces. This tendency is consistent with equation 6.9, which predicts a hotter and photon-denser environment close to the SMBH. Remarkably, the EIC process starts to dominate the electron cooling at a distance range $\mathcal{R} \lesssim 10^2$, leading to a softer electron spectrum, e.g., the blue line ($\mathcal{R} = 10$) in the right panel of Fig. 6.2, in contrast to the high-$\mathcal{R}$ cases. The black solid line in the right panel shows the electron injection function. In the low-energy band, there is no injection, e.g., $\dot{Q}_{e,\rm inj} = 0$ for $\gamma'_e \lesssim \gamma'_{e,m}$, we can analytically solve equation 6.10 and connect this segment to the $\gamma'_e > \gamma'_{e,m}$ part. Using the simplification $t'^{-1}_c \sim b\gamma'_e$, which is consistent with the EIC and synchrotron cooling rates in the left panel, we obtain

$$\mathcal{N}_{\gamma'_e} = \mathcal{N}_{\gamma'_{e,m}} \exp\left[-\frac{1}{bt'_{\rm dyn}}(\gamma'_{e,m} - \gamma'_e)\right],\ \gamma'_e \lesssim \gamma'_{e,m}, \tag{6.12}$$

where $\mathcal{N}_{\gamma'_{e,m}}$ represents the electron number distribution at $\gamma'_{e,m}$. Equation 6.12 explains the electron spectrum softening at lower values of $\mathcal{R}$ (equivalently at larger values of b).

6.4 Results

6.4.1 γ-Ray Spectra

Using the electron spectra obtained in Sec. 6.3 and following the formalism and procedures presented in [401], [403] and [363], we numerically compute the γ-ray spectra taking into account the synchrotron, SSC and EIC processes. We consider three merger-induced

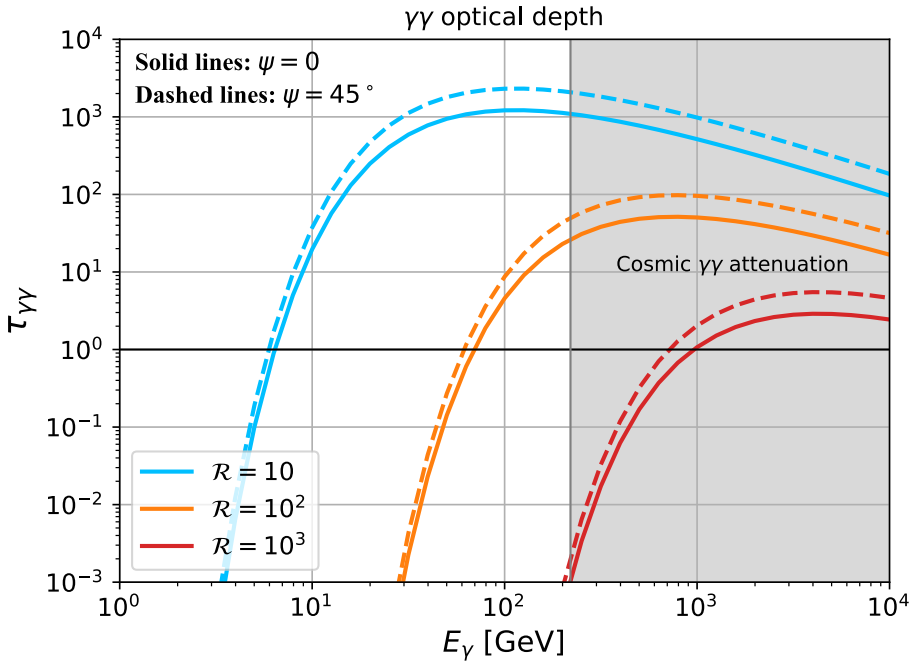

Figure 6.3. The blue ($\mathcal{R} = 10$), yellow ($\mathcal{R} = 10^2$) and red ($\mathcal{R} = 10^3$) lines are the optical depth $\tau_{\gamma\gamma}$ for $\gamma\gamma$ annihilation between γ-rays and disk photons. The solid and dashed lines correspond to the inclination $\psi = 0$ and $\psi = 45°$. The optical depth to cosmic $\gamma\gamma$ annihilation becomes greater than 1.0 in the energy range $E_\gamma \gtrsim 220$ GeV (the gray shaded area), assuming that the CBO merger is located at $z = 1.0$.

GRBs in an AGN located at redshift $z = 1$ (the equivalent luminosity distance is $d_L \simeq 6.7$ Gpc). We focus on the on-axis case and assume the CBOs' orbit planes are all aligned with the AGN disk plane, e.g., $\psi = 0$. A discussion on the influence of ψ will be given in Sec. 6.4.2.

While propagating in the jet and in the AGN disk, high-energy γ-rays will annihilate with ambient UV/IR disk photons, resulting in their attenuation and EM cascades. The optical depth for $\gamma\gamma$ annihilation depends on the photon energy in the short GRB's engine frame $\varepsilon_\gamma = \Gamma_j \varepsilon'_\gamma$, the position of the jet and the misalignment angle ψ, via

$$\tau_{\gamma\gamma}[\varepsilon_\gamma, \mathcal{R}, \psi] \approx \int_0^{H_{\rm AGN}} \frac{dy}{\cos\psi} \lambda_{\gamma\gamma}^{-1}[\varepsilon_\gamma, R_d + y\tan\psi], \tag{6.13}$$

where the reciprocal of the mean free path $\lambda_{\gamma\gamma}[R_d]$ for an isotropic disk photon field can be calculated as [420]

$$\lambda_{\gamma\gamma}^{-1}[\varepsilon_\gamma, R_d] = \frac{1}{2}\int_{-1}^{1} d\mu(1-\mu) \int d\tilde{\varepsilon}_\gamma n_{\epsilon_\gamma}^{(\rm eic)}[\tilde{\varepsilon}_\gamma]\sigma_{\gamma\gamma}[x]. \tag{6.14}$$

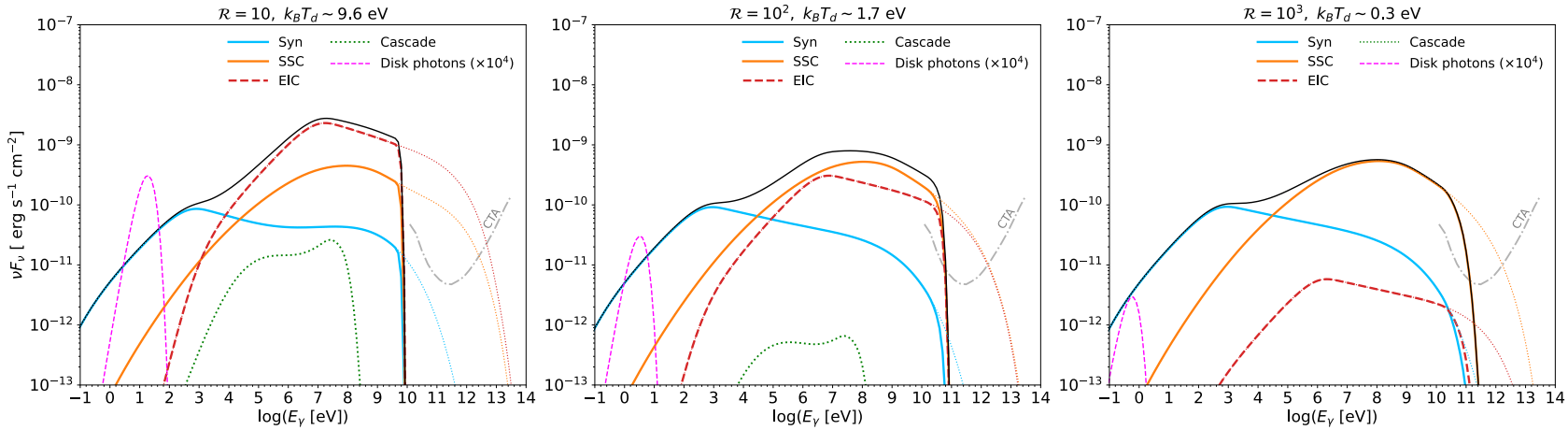

Figure 6.4. The observed γ-ray spectra from embedded short GRBs at $z = 1$ with distances $\mathcal{R} = 10$ (left panel), 10^2 (middle panel) and 10^3 (right panel) to the central SMBH. The GRB parameters used here are the fiducial parameters assumed in Sec. 6.3, e.g., $L_{j,\rm iso} = 10^{48.5}$ erg s^{-1}, $\Gamma_j = 50$, $\epsilon_B = 0.01$, and $\epsilon_e = 0.1$. The blue, yellow and red solid lines show the synchrotron, SSC and EIC emission after $\gamma\gamma$ attenuation. The dotted lines in the corresponding colors depict the unattenuated fluxes. The cascade emissions are depicted as the green lines. The magenta dashed lines show the disk target photon fluxes (multiplied by 10^4). In both cases, $\psi = 0$ is applied. The gray dash-dotted lines indicate the CTA flux sensitivity for the 10^3 s observation time.

In this expression, $x = \tilde{\varepsilon}_\gamma \varepsilon_\gamma (1-\mu)/2$ is the particle Lorentz factor in the center-of-momentum frame and $\sigma_{\gamma\gamma}$ is the $\gamma\gamma$ annihilation cross section.

Fig. 6.3 shows the optical depth in the observer's frame, where the observed energy is connected with ε_γ and ε'_γ via $E_\gamma = \varepsilon_\gamma/(1+z) = \Gamma_j \varepsilon'_\gamma/(1+z)$. The solid blue, yellow and red lines illustrates $\tau_{\gamma\gamma}$ at $\mathcal{R} = 10,\ 10^2$ and 10^3 with $\psi = 0$, whereas the dashed lines correspond to the case of an inclined jet, e.g., $\psi = 45°$. The universe becomes opaque for γ-rays produced at $z = 1$ with energies $E_\gamma \gtrsim 220$ GeV (see the gray area in Fig. 6.3) due to $\gamma\gamma$ annihilation between γ-rays and cosmic backgrounds [421] e.g. extragalactic background light (EBL) and cosmic microwave background (CMB). From Fig. 6.3, we find that γ-rays with energy $E_\gamma \gtrsim 10$ GeV are strongly suppressed due to $\gamma\gamma$ annihilation for a GRB close to the SMBH, i.e. $\mathcal{R} \simeq 10$. For a GRB at positions with a larger $\mathcal{R} \sim 10^2 - 10^3$, γ-ray photons with energy $E_\gamma \sim 100$ GeV can escape from the AGN disk.

Applying the factor $\exp(-\tau_{\gamma\gamma})$ to the γ-ray spectra, we obtain the $\gamma\gamma$-attenuated spectra for embedded GRBs at redshift $z = 1$, as shown in Fig. 6.4. In this figure, $\psi = 0$ is used. The blue solid, yellow solid, and red dashed lines respectively illustrate the synchrotron, SSC, and EIC components. The dotted lines with corresponding colors show the fluxes before $\gamma\gamma$ attenuation. The gray dash-dotted lines indicate the Cherenkov Telescope Array (CTA) flux sensitivity for the 10^3 s observation time [422]. The magenta dashed lines show the disk photon fluxes multiplied by 10^4. From the red dashed lines in Fig. 6.4, we find that a closely embedded GRB can produce brighter γ-ray emission due to the EIC enhancement. The "Compton dominance" induced by EIC enhancement can be used as the prominent feature to distinguish these embedded short GRBs from others.

The e^+/e^- pairs produced in the $\gamma\gamma$ annihilation process will induce electromagnetic cascades while diffusing and cooling down in the AGN disk via synchrotron and inverse Compton processes. Following the treatment in [423], we write down the distribution for the secondary electrons and positrons,

$$\mathcal{N}^{\rm cas}_{\gamma_e} \approx 2\mathcal{N}^{\rm ph}_{\hat{\varepsilon}_\gamma} \left(\frac{d\hat{\varepsilon}_\gamma}{d\gamma_e}\right) \left(1 - e^{-\tau_{\gamma\gamma}[\hat{\varepsilon}_\gamma, \mathcal{R}, \psi]}\right) \tag{6.15}$$

where $N^{\rm ph}_{\hat{\varepsilon}_\gamma}$ is the pre-attenuation gamma-ray number spectra (in the units of eV^{-1}) in the engine frame and $\hat{\varepsilon}_\gamma = 2\gamma_e m_e c^2$ is the energy of primary electrons. Using the cavity

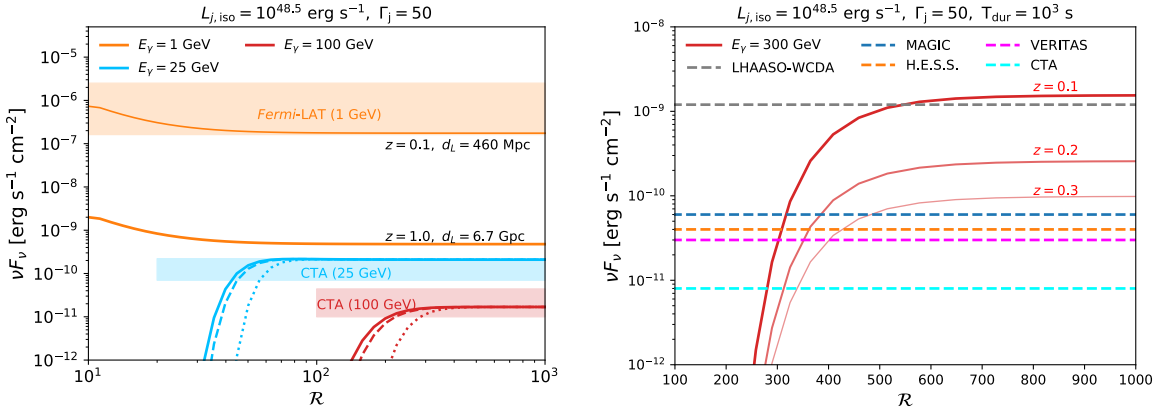

Figure 6.5. *Left panel:* γ-ray fluxes at 1 GeV (yellow lines), 25 GeV (blue lines) and 100 GeV (red lines) as functions of $\mathcal{R}$. The thick lines are obtained with $L_{j,\rm iso} = 10^{48.5}$ erg s^{-1} and $z = 1.0$, whereas a closer short GRB at $z = 0.1$ is considered for the thin yellow line. The point-source performance for *Fermi*-LAT and CTA at corresponding energies are shown as the yellow, blue and red areas, respectively. The upper and lower bounds show the sensitivities for the observation time $T_{\rm dur} = 10^2$ s and $T_{\rm dur} = 10^3$ s. *Right panel:* The red solid lines from thick to thin show the $\mathcal{R}$-dependence of 300 GeV γ-ray fluxes from the embedded short GRBs at $z = 0.1$, 0.2, and 0.3. The horizontal dashed lines from top to bottom correspond the the sensitivities of LHAASO-WCDA, MAGIC, H.E.S.S., VERITAS, and CTA.

magnetic field

$$B_{\rm cav} \approx (2\epsilon_B \eta_w \dot{M}_{\rm CBO} v_w / H_{\rm AGN}^2)^{1/2} \simeq 98\ \epsilon_{B,-2}^{1/2} \eta_{w,-0.5}^{1/2} h_{\rm AGN.-2}^{-1} \eta_{\rm CBO,-1}^{1/2} \mathcal{R}_2^{-1} \dot{m}_\star^{1/2} M_{\star,8}^{-1/2} v_{w,9}^{1/2}\ {\rm G}, \tag{6.16}$$

we numerically calculate the cascade emission. The green dotted lines in Fig. 6.4 show the cascade emission. Comparing to the beamed emission produced in the jet, the cascade emission is subdominant for $\mathcal{R} \gtrsim 100$ and typically peaks at a lower energy ~ 100 MeV. We find that the cascade flux drops dramatically as $\mathcal{R}$ increases, which is consistent with the $\mathcal{R}$-dependence of the $\gamma\gamma$ optical depth in Fig. 6.3. When the disk becomes transparent to the γ-ray photons, the e^-/e^+ pair production is suspended and the cascade emission is strongly suppressed. Typically, we need to solve the time-dependent equations to obtain the secondary electron/position distributions and the cascade spectrum. Our approach can provide a good estimation since these secondary particles cools down very fast, e.g., $t_{e,c}^{\rm cas} \lesssim 10$ s.

6.4.2 Detectability with Fermi-LAT and VHE γ-Ray Facilities

It is useful to compare the expected γ-ray fluxes in the extended emission phase against the sensitivities of current and future facilities, such as *Fermi*-LAT, MAGIC, H.E.S.S., VERITAS, CTA, and LHAASO-WCDA, and discuss how the parameters $\mathcal{R}$ and ψ influence the results.

Observationally, a significant fraction of short GRBs exhibit 'long-lasting' extended or plateau emission peaking in X-ray bands [424–427] with the duration $T_{\rm dur} \sim 10^2 - 10^5$ s, following the prompt phase where 90% of the kinetic energy is dissipated in ~ 2 seconds, e.g., $T_{90} \lesssim 2$ s. Such prolonged emission may originate from the continuous energy injection by the accreting black holes formed after the merger or the fast rotating magnetars [428–434]. Considering a prolonged γ-ray emission of luminosity $L_{j,\rm iso} = 10^{48.5}$ erg s^{-1} and the corresponding duration in the observer's frame $T_{\rm dur} \sim 10^2$ s $- 10^3$ s, we show the integral sensitivities within $T_{\rm dur}$ for *Fermi*-LAT[1] and CTA [422] at the $E_\gamma =$ 1 GeV (yellow area), 25 GeV (blue area) and 100 GeV (red area) in the left panel of Fig. 6.5. The upper and lower bounds of each shaded area demonstrate the performances for the detectors given the observation time $T_{\rm dur} = 10^2$ s and $T_{\rm dur} = 10^3$ s, respectively. We plot also the 1 GeV (yellow lines), 25 GeV (blue lines) and 100 GeV (red lines) fluxes as functions of $\mathcal{R}$ in the left panel of Fig. 6.5. The solid lines correspond to the $\psi = 0$ case, whereas the dashed and dash-dotted lines depict the $\psi = 45^\circ$ and $\psi = 75^\circ$ cases. The thick lines are for the GRBs at $z = 1$, while the thin yellow line shows the 1 GeV fluxes for a closer GRB at $z = 0.1$ ($d_L \simeq 460$ Mpc).

The influence of disk photons is encoded in the shapes of the yellow, blue and red curves. The 1 GeV flux decreases to a flat level as $\mathcal{R}$ increases because the EIC component gradually becomes less important as the CBO is moved to a cooler outer region. In the ranges $\mathcal{R} \lesssim 50$ and $\mathcal{R} \lesssim 300$, the $\gamma\gamma$ attenuation caused by dense disk photons suppresses the 25 GeV and 100 GeV emission, respectively. Since the $\gamma\gamma$ annihilation is negligible for 1 GeV photons even if the CBO is very close to the SMBH (see the blue lines in Fig. 6.3), we expect that the flux does not depend on ψ. On the other hand, the 25 GeV and 100 GeV fluxes decrease as ψ approaches $\psi_c \simeq 85.6^\circ$.

From the left panel of Fig. 6.5, we find that CTA will be capable of detecting 25 GeV and 100 GeV γ-rays up to $z = 1$ if an embedded short GRB is appropriately distant from the SMBH, e.g., $\mathcal{R} \gtrsim 40$ for 25 GeV γ-rays and $\mathcal{R} \gtrsim 200$ for 100 GeV γ-rays. By contrast, it is challenging for *Fermi*-LAT to detect the 1 GeV photons from sources located at

[1]The *Fermi*-LAT sensitivity can be found in https://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm

$z = 1$ via point source search within the duration $T_{\rm dur} \sim 10^3$ s. For the short GRBs embedded in AGN disks, we would require a nearby CBO merger ($d_L \lesssim 460$ Mpc) at the position with the distance greater than $40R_{\rm S}$ ($\mathcal{R} \gtrsim 40$) to the central SMBH in order to be detected simultaneously by CTA and *Fermi*-LAT.

MAGIC, H.E.S.S., and VERITAS are current ground Imaging Atmospheric Cherenkov Telescopes with very good performance in the energy range 150 GeV to 30 TeV. LHAASO is a new generation multi-component instrument and LHAASO-WCDA is operated in the energy range ∼300 GeV to 10 TeV. We present the $\mathcal{R}$-dependence of 300 GeV γ-ray fluxes at $z = 0.1$, 0.2, and 0.3 (the red solid lines, from thick to thin) in the right panel of Fig. 6.5. The horizontal dashed lines from top to bottom corresponds to the flux sensitivities of LHAASO-WCDA [435], MAGIC [436], H.E.S.S. [437], VERITAS[2], and CTA for $T_{\rm dur} = 10^3$ s and $\psi = 0$. At 300 GeV, the sensitivity of LHAASO-WCDA is $\sim 10^{-9}$ erg s^{-1} cm^{-2} in 10^3 s observation. The nearby embedded GRBs with redshift $z < 0.1$ can be observed. MAGIC, H.E.S.S., VERITAS and CTA can detect 300 GeV photons from embedded GRBs upto redshift $z = 0.3$ if $\mathcal{R} \gtrsim 500$ is satisfied. For the sources with farther distance, the Universe could be opaque to VHE γ-rays.

6.4.3 Prompt Emission

As for the prompt emission, besides the cutoff with energy $\gtrsim 100$ GeV caused by the $\gamma\gamma$ absorption in the AGN disk, we found that there may be no significant difference between short GRBs embedded in AGN disks and other short GRBs. The reason is that, given a higher isotropic luminosity $L_{j,\rm iso}^{\rm prompt} = 10^{51}$ erg s^{-1} and a higher Lorentz factor $\Gamma_j^{\rm prompt} = 200$ ($\Gamma_j^{\rm prompt} = 100$) in the prompt emission phase of $T_{90} = 1$ s, the EIC emission is subdominant (comparable) compared to the synchrotron/SSC components. Using the parameters in the prompt emission phase, we estimate photon flux in the energy range 50 - 300 keV,

$$F_{\nu,50-300\ \rm keV}^{\rm prompt} \simeq 1.9\ (1+z) d_{L,28}^{-2}\ \rm ph\ s^{-1}\ cm^{-2}. \tag{6.17}$$

Noting that the onboard trigger threshold of the *Fermi* Gamma-Ray Burst Monitor (*Fermi*-GBM) is $\sim$ 0.7 ph s^{-1} cm^{-2} [438], it can detect the prompt emission and localize the short GRB. At 10 GeV, the flux of the prompt emission is $\nu F_{\nu,10\ \rm GeV}^{\rm prompt} \sim 2\times10^{-6}\ (1+z) d_{L,28}^{-2}$ erg s^{-1} cm^{-2}, implying the possible detection of the embedded GRBs

[2]The differential sensitivity of VERITAS can be found in https://veritas.sao.arizona.edu/about-veritas/veritas-specifications

at $z \sim 0.5 - 1$ with the High Altitude Water Cherenkov (HAWC) observatory [439]. If the short GRB is GRB 090510-like, e.g., $L_{j,\rm iso}^{\rm prompt} \gtrsim 10^{53}$ erg s^{-1}, *Fermi*-LAT would also be able to see γ-ray photons upto ~ 30 GeV in the prompt emission phase [440]. Above all, the prompt emission diagnosis can provide valuable information for the follow-up observations of extended emissions.

6.5 Summary and Discussion

We studied γ-ray emission from short GRBs embedded in AGN disks and showed that successful jets are expected from these, since the CBOs in the disks are highly super-Eddington accretors and can produce low-density cavities around the CBO via powerful outflows. Our work demonstrates that the AGN disks influence the γ-ray emission mainly in two ways, namely, via the EIC enhancement and $\gamma\gamma$ attenuation, depending on the distance to the SMBH and the inclination ψ. If a CBO merger occurs very close to the SMBH, e.g., $\mathcal{R} \sim 10 - 40$, the dense disk photon field will lead to a luminous EIC component in the GeV band and a firm cutoff at $E_\gamma \simeq 10$ GeV. On the other hand, the SSC process dominates the GeV emission for CBO mergers at $\mathcal{R} \gtrsim 100$, and the disk gradually becomes transparent for 10-100 GeV photons unless the GRB jet is entirely buried inside the AGN disk, e.g., $\psi \gtrsim \psi_c \simeq 85.6°$. Considering the ratio of the peak flux of the inverse Compton component to the synchrotron peak flux and the cutoff energy, we may be able to distinguish the short GRBs embedded in AGN disks from other types of isolated short GRBs [107, 108]. To identify the embedded short GRBs, we can utilize these two signatures, "Compton dominance" and $\gamma\gamma$ annihilation cutoff. Such spectral information can also be used to determine the parameters of the short GRB - AGN disk system such as T_d, $\mathcal{R}$ and ψ. According to the simulations of compact binary formations in AGN disks, it is reasonable to expect the embedded short GRBs to occur in the region $\mathcal{R} \gtrsim 40 - 100$ [115, 124]. The detection of these short GRBs can, in return, be used to test current AGN-assisted CBO formation theories and constrain the CBO distributions in AGN disks.

Since approximately $f_{\rm EE} \sim 1/4 - 1/2$ [431] of *Swift* short GRBs are accompanied by extend emission, we investigated the detectability of GRBs in the AGN disk for CTA and *Fermi*-LAT considering a jet of luminosity $L_{j,\rm iso} = 10^{48.5}$ erg s^{-1} lasting for $T_{\rm dur} \sim 10^2 - 10^3$ s. From now on, we discuss the detection perspectives of the extended emissions with $T_{\rm dur} = 10^2 - 10^3$ s, $L_{\rm j,iso} = 10^{48.5}$ erg s^{-1} cm^{-2}, and $\Gamma_j = 50$. For the embedded short GRBs within $z = 1.0$, CTA will be able to detect the γ-rays in the

energy range $E_\gamma \sim 25 - 100$ GeV if the requirements $\mathcal{R} \gtrsim \mathcal{R}_c$ and $\psi \lesssim \psi_c$ are satisfied, where $\mathcal{R}_c \sim 40 - 100$ is the critical distance defined by $\tau_{\gamma\gamma}[(1+z)E_\gamma, \mathcal{R}_c, \psi] = 1$. To estimate the CTA detection rate, we use $f_\mathcal{R}$ and $f_\psi \sim 1$ to represent the fractions of embedded short GRBs that meet the conditions $\mathcal{R} \gtrsim \mathcal{R}_c$ and $\psi \lesssim \psi_c$, respectively. Taking into account both NS-NS and NS-BH mergers, [441] estimated the occurance rate of short GRB in AGN disks at $z < 1$, $\dot{R}_{\rm SGRB,AGN} \sim (300 - 2 \times 10^4) f_{\rm AGN,-1}$ yr^{-1}, where $f_{\rm AGN} \sim 0.1$ is the fraction of BH-BH mergers. We estimate the CTA detection rate of the on-axis prolonged γ-ray emission from short GRBs embedded in AGN disks via $\dot{R}_{\rm CTA} \sim f_{\rm CTA} f_b f_{\rm EE} f_\mathcal{R} f_\psi \dot{R}_{\rm GRB,AGN} \sim (0.2 - 22)\ f_\mathcal{R} \theta_{j,-1}^2 f_{\rm AGN,-1}$ yr^{-1}, where $f_{\rm CTA} \sim 0.3 - 0.5$ is the CTA detection efficiency defined as the ratio of detectable events to events that can be followed up by CTA [442], $f_b = (\theta_j + 1/\Gamma_j)^2/2 \sim \theta_j^2/2$ is the beaming factor and $\theta_j \sim 0.1$ is the jet opening angle. Despite the large uncertainty in the CTA detection rate, we estimate that it is feasible for CTA to detect the prolonged γ-ray emission from short GRBs embedded in AGN disks in the time scale of one year.

We now discuss the implications to multi-messenger analyses with GWs and γ-rays. [443] estimated that the merger rate of binary black holes (BBHs) embedded in AGN disks within the advanced Laser Interferometer Gravitational-wave Observatory's (aLIGO's) horizon, e.g., $D_h \simeq 450$ Mpc, could be $\dot{R}_{\rm L,BBH} \sim 20$ yr^{-1}. Implementing the ratio of the cumulative NS-BH and NS-NS merger rates to the BBH merger rate in the AGN channel, $f_{\rm L,CBO/BBH} = (\dot{R}_{\rm L,NS-NS} + \dot{R}_{\rm L,NS-BH})/\dot{R}_{\rm L,BBH} \sim 0.1 - 7.0$ [441], we estimate the occurrence rate of on-axis short GRBs with extended emission originating from LIGO-detectable CBO mergers in the AGN channel,

$$\begin{aligned} \dot{R}^{(L)}_{\rm SGRB-AGN} &= f_{\rm EE} f_b f_{\rm L,CBO/BBH} \dot{R}_{\rm L,BBH} \\ &\sim (2.5 \times 10^{-3} - 0.35)\ \theta_{j,-1}^2\ {\rm yr}^{-1}. \end{aligned} \tag{6.18}$$

The physical meaning of this equation is that among all detectable mergers within LIGO's horizon, MAGIC, H.E.S.S., VERITAS, CTA, and LHAASO-WCDA can observe $2.5 \times 10^{-3} - 0.35$ short GRBs with extended γ-ray emission each year. In the optimistic case, it is possible to detect the on-axis extended emission simultaneously with GWs originated from CBO mergers embedded in AGN disks in one decade.

We note also that, while this is not the subject of the present work, the model predicts that short GRBs from CBO mergers are efficient neutrino emitters. Our model does not require choked jets, unlike [122, 444]. The CRs accelerated in the successful jet can efficiently interact with disk photons and produce high-energy neutrinos via the photomeson production process. Using equations 8 and 9 of [445] and Fig. 6.3 of this

work, the photomeson optical depth is $f_{p\gamma} \sim 1$ for $\mathcal{R} \sim 10$ and $f_{p\gamma} \sim 0.1$ for $\mathcal{R} \sim 100$. High-energy neutrinos are expected in the PeV range, and they will make additional contribution to those predicted by [446]. The enhancement is more prominent for prompt neutrino emission, because the efficiency is low for usual short GRBs.

In conclusion, future multi-messenger analyses of AGN short GRBs can provide unprecedented insights for understanding the formation and evolution of CBOs inside the AGN disks as well as on the origin of their high-energy emission.

Acknowledgements

We thank B. Theodore Zhang, Mukul Bhattacharya, Zsuzsa Márka and Szabolcs Márka for fruitful discussions. C.C.Y. and P.M. acknowledge support from the Eberly Foundation. The work of K.M. is supported by the NSF Grant No. AST-1908689, No. AST-2108466 and No. AST-2108467, and KAKENHI No. 20H01901 and No. 20H05852. A.P. is supported by the European Research Council via ERC consolidating grant 773062 (acronym O.M.J.). I.B. acknowledges the support of NSF under awards PHY-1911796 and PHY-2110060 and the Alfred P. Sloan Foundation.

Disclaimer

The findings and conclusions do not necessarily reflect the view of the funding agencies.

Chapter 7
Complementarity of Stacking and Multiplet Constraints on the Blazar Contribution to the Cumulative Diffuse Neutrino Flux

Note: The material in this Chapter is based on my paper [447], with co-authors Kohta Murase, and Peter Mészáros.

7.1 Introduction

Since the initial detection of high-energy astrophysical neutrinos by the IceCube Neutrino Observatory [264, 265], a cumulative flux of astrophysical neutrinos in the energy range from $\sim$ 10 TeV to several PeV has been unveiled and measured to a higher precision [32, 189, 266]. The isotropic distribution of the cumulative flux as well as the background-only results from recent searches for point-like sources and multi-messenger analyses support an extragalactic origin of these neutrinos [133, 448, 449]. Up to now, however, the main origin of the cumulative neutrinos still remains unknown.

The flavor ratio measured at Earth, $(\nu_e : \nu_\mu : \nu_\tau) \approx (1 : 1 : 1)$, is consistent with the prediction from the long-distance oscillations of neutrinos produced through pion decays [450], which provides one common framework for the astrophysical models dedicated to explain the cumulative neutrino flux. Many candidates have been proposed and studied [451, 452]. Among these candidates, blazars, which are known as a subclass of AGNs with a relativistic jet pointing nearly towards the Earth [453, 454], have

been frequently considered as promising ultra-high-energy cosmic-ray (CR) accelerators and high-energy neutrino emitters [58, 455–457]. Recently, the IceCube collaboration announced the spatial and temporal coincidence between a muon track neutrino event IceCube170922A and a blazar TXS 0506+056 [125] at the significance $\sim 3\sigma$. Intuitively, if this association is physical, the intimate link between this IceCube neutrino event and the blazar may favor blazars as the main sources of the cumulative neutrino flux, but this is not the case [8].

The maximum likelihood stacking searches for cumulative neutrino flux from the second *Fermi*-LAT AGN catalog (2LAC) as well as the point-source searches using the IceCube muon track events and blazars in *Fermi*-LAT 3LAC have independently shown that *Fermi*-LAT-resolved blazars only contribute a small portion of the IceCube cumulative neutrino flux [126–128] and the hadronic models of blazar activity are strongly constrained [129], if the specific correlation $L_\nu \propto L_{\rm ph}$ is assumed as a prior. [130] evaluated the contribution of unresolved sources, and showed that the blazar contribution to the cumulative neutrino flux is constrained unless one makes an ad hoc assumption that lower-luminosity blazars entrain a larger amount of CRs.

Here we argue that, in addition to the stacking analysis, the absence of clustering in high-energy neutrino events, i.e., neutrino multiplets and auto-correleation, can also provide relevant constraints on various classes of proposed sources as the dominant origin of the cumulative neutrino flux [131–136]. The constraints are sensitive to the redshift evolution of the sources, which are especially powerful for weakly or non-evolving sources such as BL Lac objects [8,131]. But the limits are weaker for rapidly evolving sources such as FSRQs, which could significantly alleviate the constraints, as remarked by [8]. [137] studied the constraints on evolving blazar populations and confirmed that fast evolving sources (e.g., $\xi_z = 5.0$) may indeed relax the neutrino multiplet limits.

In this work, we consider the "joint" implications of these independent analyses for the global blazar population and extend the constraints to a common case where a generic relationship between neutrino and gamma-ray luminosities, e.g., $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$, is presumed, which is more general than what has been previously considered in such analyses. Physically, the correlation between L_ν and $L_{\rm ph}$ is determined by the interactions between particles and radiation fields inside the sources. Most of physically reasonable models developed on the basis of photohadronic (e.g., $p\gamma$) interactions predict $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$ with indices of $1.0 \lesssim \gamma_{\rm lw} \lesssim 2.0$ [8, 54, 57, 58, 131, 458–461]. The index $\gamma_{\rm lw}$ characterizes the source models and may deviate from this fiducial range for models with increasing complexity. Motivated by this, we treat $\gamma_{\rm lw}$ as a free parameter and attempt

to reveal the $\gamma_{\rm lw}$-dependence of the upper limits on all-blazar contributions. In addition, a new feature of our analysis is that we also consider the effect of *Fermi*-unresolved blazars. One caveat is that, in this study, we assume all sources are equal and emit steadily with a single power-law spectrum. Prior to the the IceCube-170922A alert, IceCube collaboration has found a neutrino excess from the direction of TXS 0506+056 during a 158-day time window in 2014-2015 [267], which reveals the the transient nature of the neutrino emission. We need to keep in mind that the multiplet limits are stronger for flaring sources [8]. The stacking limits are also applicable to time-averaged emission of the flaring sources, as long as the scaling between neutrino and gamma-ray luminosities hold [8].

In the first part (Sec. 7.2), we calculate the ratio of neutrino fluxes from *Fermi*-LAT-resolved blazars and all blazars (including both resolved and unresolved contributions). Combining this ratio with the existing constraints on *Fermi*-LAT-resolved blazars, we estimate the upper limits for all-blazar contributions. The multiplet constraints are given in the second part (Sec. 7.3) where we also derive the effective number densities $n_0^{\rm eff}(\gamma_{\rm lw})$ and the redshift evolution factor $\xi_z(\gamma_{\rm lw})$ for blazars and the subclasses, FSRQs and BL Lacs. In either case, we use the blazar gamma-ray luminosity functions provided by [462–464] to reconstruct the neutrino luminosity density. In Sec. 7.4 we conclude with a discussion.

7.2 Implications of Stacking Limits

Given the differential density of blazars as a function of rest-frame 100 MeV-100 GeV luminosity $L_{\rm ph}$, redshift z and photon index Γ defined by the gamma-ray flux $F \propto \varepsilon_{\rm ph}^{-\Gamma}$,

$$\frac{d^3N_{\rm bl}}{dL_{\rm ph}dzd\Gamma} = \phi_{\rm bl}(L_{\rm ph}, z)\frac{dP_{\rm bl}}{d\Gamma}\frac{dV}{dz}, \tag{7.1}$$

where the subscript "bl" represents blazars considered in the calculation, $\phi_{\rm bl}(L_{\rm ph}, \Gamma) = d^2N_{\rm bl}/dL_{\rm ph}dV$ is the luminosity function and $dP_{\rm bl}/d\Gamma$ is the probability distribution of spectral index Γ, we can directly write down the (differential) luminosity density of neutrinos from *Fermi*-LAT-resolved blazars at redshift z,

$$\begin{aligned}\varepsilon_\nu Q_{\varepsilon_\nu}^{\rm (bl,R)}(z, \gamma_{\rm lw}) = \int_{L_{\rm ph,th}}^{L_{\rm ph,max}} \int_{\Gamma_{\rm min}}^{\Gamma_{\rm max}} \mathcal{C}^{-1}\phi_{\rm bl}(L_{\rm ph}, z)L_\nu(L_{\rm ph}) \\ \times \frac{dP_{\rm bl}}{d\Gamma}d\Gamma dL_{\rm ph}\end{aligned} \tag{7.2}$$

where $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$ is the neutrino luminosity, $L_{\rm ph,max}$ is a fixed upper limit of blazar luminosity and the lower limit $L_{\rm ph,th}(L_{\rm ph}, z, \Gamma)$ is determined by the *Fermi* LAT threshold flux $F_{100,\rm th}$ in the energy range of 100 MeV – 100 GeV. In this equation, $\mathcal{C}$ is the normalization coefficient determined by $\varepsilon_{\rm CR,max}$ and $\varepsilon_{\rm CR,min}$, the maximum and minimum energy that CRs in blazars can achieve. Since we aim to estimate the neutrino flux from a general luminosity relationship, $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$, and the physics may be unknown for a general $\gamma_{\rm lw}$, we do not try to provide the details of the gamma-ray and neutrino radiation processes. In this work, we assume that the maximum CR energy is the same for all blazars, as is the normalization factor once the spectral index s of the IceCube neutrino flux is specified.

Here, we present one method to rewrite the integrals in equation 7.2 by incorporating the *Fermi*-LAT detection efficiency. For a blazar at redshift z with the luminosity $L_{\rm ph} \propto \int_{\varepsilon_{\rm min}}^{\varepsilon_{\rm max}} F(\varepsilon)\varepsilon d\varepsilon$, where $\varepsilon_{\rm max} = 100(1+z)$ GeV and $\varepsilon_{\rm min} = 100(1+z)$ MeV, and the photon index Γ, the integrated photon flux at earth can be written as

$$\begin{aligned} F_{100}(L_{\rm ph}, z, \Gamma) &= \int_{\varepsilon_{\rm min}}^{\varepsilon_{\rm max}} F(\varepsilon) d\varepsilon \\ &= \frac{L_{\rm ph}}{4\pi d_L^2(z)} \times \begin{cases} \ln\left(\frac{\varepsilon_{\rm max}}{\varepsilon_{\rm min}}\right) \frac{1}{\varepsilon_{\rm max}-\varepsilon_{\rm min}} & \Gamma = 1 \\ \frac{\varepsilon_{\rm max}-\varepsilon_{\rm min}}{\varepsilon_{\rm max}\varepsilon_{\rm min}\ln\left(\frac{\varepsilon_{\rm max}}{\varepsilon_{\rm min}}\right)} & \Gamma = 2 \\ \frac{2-\Gamma}{1-\Gamma}\frac{\varepsilon_{\rm max}^{1-\Gamma}-\varepsilon_{\rm min}^{1-\Gamma}}{\varepsilon_{\rm max}^{2-\Gamma}-\varepsilon_{\rm min}^{2-\Gamma}} & \Gamma \neq 1, 2, \end{cases} \end{aligned} \tag{7.3}$$

where d_L is the luminosity distance between the blazar and the detector. Then the lower limit of the integral in equation 7.2 can be obtained by requiring $F_{100}(L_{\rm ph,th}, z, \Gamma) = F_{100,\rm th}$. Alternatively, thanks to the *Fermi*-LAT detection efficiency $\epsilon(F_{100})$ provided by [465], we can simplify equation 7.2 by using the equivalent detection efficiency $\epsilon(L_{\rm ph}, z, \Gamma) = \epsilon(F_{100})$,

$$\begin{aligned} \varepsilon_\nu Q_{\varepsilon_\nu}^{\rm (bl,R)}(z, \gamma_{\rm lw}) = \int_{L_{\rm ph,min}}^{L_{\rm ph,max}} \int_{\Gamma_{\rm min}}^{\Gamma_{\rm max}} \mathcal{C}^{-1} \phi_{\rm bl}(L_{\rm ph}, z) L_\nu(L_{\rm ph}) \\ \times \epsilon(L_{\rm ph}, z, \Gamma) \frac{dP_{\rm bl}}{d\Gamma} d\Gamma dL_{\rm ph}, \end{aligned} \tag{7.4}$$

where the lower limit $L_{\rm ph,min}$ reduces to a constant and represents the minimal luminosity of blazars that are considered in this work. To eliminate the instrumental selection effect produced by the low detection efficiency for dimmer blazars and to take all blazars into account, we replace the $L_{\rm ph,th}$ in equation 7.2 by $L_{\rm ph,min}$, which yields the neutrino

luminosity density from all blazars $\varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,all)}(z,\gamma_{\rm lw})$, which can be written explicitly as

$$\begin{aligned}\varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,all)}(z,\gamma_{\rm lw}) = &\int_{L_{\rm ph,min}}^{L_{\rm ph,max}}\int_{\Gamma_{\rm min}}^{\Gamma_{\rm max}} \mathcal{C}^{-1}\phi_{\rm bl}(L_{\rm ph},z)L_\nu(L_{\rm ph}) \\ &\times \frac{dP_{\rm bl}}{d\Gamma}d\Gamma dL_{\rm ph}.\end{aligned} \tag{7.5}$$

Meanwhile, using the LFs for luminosity-dependent density evolution (LDDE) models and parameters provided by [463, 464], we successfully reproduced the redshift evolution of FSRQ and BL Lac luminosity densities illustrated in the Figure 6 of [464]. At this stage, during the integration of $L_{\rm ph}$, we set the maximum and minimum luminosities to be 10^{50} erg s^{-1} and 10^{40} erg s^{-1}, respectively. We also found that the results are consistent with the uncertainties in [464] when the limits of the integration were varied by one or two orders of magnitude. Another thing that we need to keep in mind is that we assume the *Fermi*-LAT-unresolved blazars share the identical LFs with the resolved ones. [466] pointed that the index distributions for different blazar classes both for the detected ones and undetected ones are slightly different: the photon spectra of newly-detected FSRQs are slightly softer than the 2LAC ones ($\Delta\Gamma < 0.1$) while in contrast there is no significant spectral difference between the two sets of BL Lacs. For the completeness, we also consider a deviation, e.g., 0.2, of the photon spectral index from the best-fit values provided by [462–464]. Such a test reveals that the resulting $\mathcal{F}(\gamma_{\rm lw})$ remains almost unchanged under a slight derivation of Γ.

Assuming the neutrino spectra from all blazars have the similar power-law form, e.g., $\varepsilon_\nu^2\Phi_{\varepsilon_\nu} \propto \varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,R/all)} \propto \varepsilon_\nu^{2-s}$, and using the comoving neutrino luminosities $\varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,all)}(z,\gamma_{\rm lw})$ and $\varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,R)}(z,\gamma_{\rm lw})$, the all-flavor neutrino fluxes from *Fermi*-LAT-resolved and all blazars at earth are expected to be

$$E_\nu^2\Phi_{E_\nu}^{(\rm bl,R/all)}(\gamma_{\rm lw}) = \frac{c}{4\pi}\int dz\ \frac{\varepsilon_\nu Q_{\varepsilon_\nu}^{(\rm bl,R/all)}(z,\gamma_{\rm lw})}{(1+z)}\left|\frac{dt}{dz}\right|, \tag{7.6}$$

where $\varepsilon_\nu = (1+z)E_\nu$. Hence, we can write down the fraction of *Fermi*-LAT-resolved blazars to the cumulative neutrino flux in a simple way that depends only on $\gamma_{\rm lw}$,

$$\mathcal{F}(\gamma_{\rm lw}) = \frac{E_\nu^2\Phi_{E_\nu}^{(\rm bl,R)}(\gamma_{\rm lw})}{E_\nu^2\Phi_{E_\nu}^{(\rm bl,all)}(\gamma_{\rm lw})}. \tag{7.7}$$

[462] presented the best-fit parameters in the blazar luminosity functions $\phi_{\rm bl}$, which enables us to compute $\mathcal{F}(\gamma_{\rm lw})$. Since the redshift correction to the energies leads to one

extra term $(1+z)^{2-s}$ to the integrand in equation 7.6 and another factor $(1+z)^{-1}$ to the integrated flux in equation 7.3, we conclude that, as a consequence, low-redshift blazars become more important when $s = 2.5$, in comparison with the $s = 2$ case. Therefore, considering nearby blazars are easier to be detected, a steeper neutrino spectrum predicts a larger $\mathcal{F}(\gamma_{\rm lw})$, which is confirmed by the thin lines in Fig. 7.1. Moreover, noting that the selection of the minimum and maximum luminosities, e.g., $L_{\rm ph,min}$ and $L_{\rm ph,max}$ of a blazar is arbitrary, we tested the reliability of $\mathcal{F}(\gamma_{\rm lw})$ by varying the integral limits and found that the results are not sensitive to $L_{\rm ph,max}$ and $\mathcal{F}(\gamma_{\rm lw})$ does not change dramatically in the range $\gamma_{\rm lw} \lesssim 1.0$ as $L_{\rm ph,min}$ increases from 10^{41} erg s^{-1} to 10^{43} erg s^{-1}, as shown in Fig. 7.1. Intuitively, a lower $L_{\rm ph,min}$ implies that more low-luminosity blazars in the sample are less likely to be detected. Also, for a weaker luminosity dependance ($\gamma_{\rm lw} \lesssim 1.0$), the low-luminosity blazars dominate the luminosity density due to the large population. The combined effect is that $\mathcal{F}(\gamma_{\rm lw})$ decreases in the range $\gamma_{\rm lw} \lesssim 1.0$. Remarkably, from Fig. 7.1, we can conclude that the contribution from *Fermi*-LAT-resolved blazars is nearly the same as the neutrino flux from all blazars when $\gamma_{\rm lw}$ is larger than 1.0. The reason is that, assuming a stronger luminosity dependance (on other words, a higher $\gamma_{\rm lw}$), the brighter blazars become increasingly important. These high-luminosity blazars have a higher chance to be detected and in this case the neutrinos luminosity densities from *Fermi*-LAT-resolved blazars and all blazars are comparable.

To compute the upper limit of cumulative neutrino flux from all blazars, we use the existing constraints, $E_\nu^2\Phi_{E_\nu}^{\rm (2LAC,stacking)}$ and $E_\nu^2\Phi_{E_\nu}^{\rm (3LAC,stacking)}$, from blazar stacking analyses and point-source searches [126, 128], which are based on *Fermi*-LAT 2LAC and 3LAC blazars. Combining these existing limits with the fraction of the neutrino flux from *Fermi*-LAT-resolved blazars, we estimate the upper limits of all-blazar contributions from *Fermi*-LAT 2LAC and 3LAC analysis,

$$E_\nu^2\Phi_{E_\nu}^{\rm (2LAC/3LAC)} = \frac{E_\nu^2\Phi_{E_\nu}^{\rm (2LAC/3LAC,stacking)}}{\mathcal{F}(\gamma_{\rm lw})}. \tag{7.8}$$

The stacking results themselves have some model dependence. Here, to obtain conservative limits, we adopt the results based on the equal flux weighting for $E_\nu^2\Phi_{E_\nu}^{\rm (2LAC/3LAC,stacking)}$. In general this gives conservative limits, and the luminosity weighting improves the constraints. We will see that, even in this most conservative case, the combined constraints of stacking and multiplet analysis are stringent.

Fig. 7.2 illustrates the upper limits for the all-blazar neutrino flux from *Fermi*-LAT 2LAC and *Fermi*-LAT 3LAC analysis. We show all-flavor neutrino fluxes for

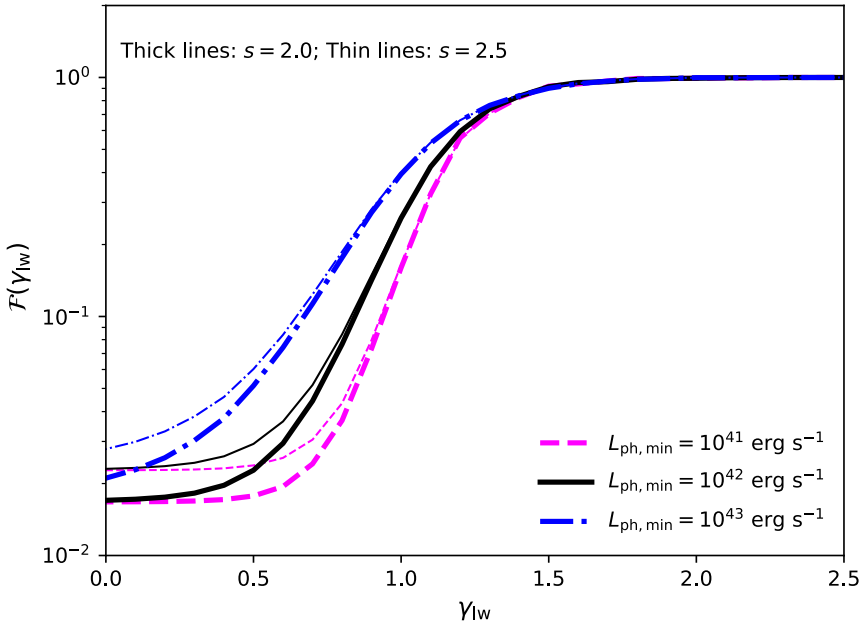

Figure 7.1. The fraction of *Fermi*-LAT-resolved blazars in the cumulative neutrino flux, $\mathcal{F}(\gamma_{\rm lw})$. The thick and thin lines are calculated for the neutrino spectral indices $s = 2.0$ and $s = 2.5$. The blue dashed, black solid and red dash-dotted lines correspond to the minimum luminosities $L_{\rm ph,min} = 10^{41}$ erg s^{-1}, 10^{42} erg s^{-1} and 10^{43} erg s^{-1}, respectively. The upper limit is fixed to be $L_{\rm ph,max} = 10^{50}$ erg s^{-1}.

all curves and data points in this figure. In the left panel, we assume $s = 2$ for the neutrino spectrum. In this case, the stacking analysis of *Fermi*-LAT-2LAC blazars gives $1.2\times10^{-8} \lesssim E_\nu^2\Phi_{E_\nu}^{\rm (2LAC,stacking)} \lesssim 1.6\times10^{-8}$ (in the unit of GeV cm^{-2} s^{-1} sr^{-1}, hereafter). The corresponding upper limits for all blazars calculated using equation 7.8 are illustrated as the magenta area. The green area in the left panel shows the constraints derived from *Fermi*−LAT 3LAC analysis which predicts $8.0\times10^{-9} \lesssim E_\nu^2\Phi_{E_\nu}^{\rm (3LAC,stacking)} \lesssim 1.4\times10^{-8}$. For the illustration purpose, we include the IceCube all-flavor neutrino flux $4.8\times10^{-8} \lesssim E_\nu^2\Phi_{E_\nu}^{\rm (IC)} \lesssim 8.4\times10^{-8}$ in Fig. 7.2 (the cyan area). To avoid underestimating the upper limits due to the uncertainties of the existing results, we introduced a 50% uncertainty to the constraints derived from stacking analysis, which broadens the areas in the left panel of Fig. 7.2. The right panel shows the energy-dependent upper limits for an $\varepsilon_\nu^{-2.5}$ neutrino spectrum. The solid lines are obtained by assuming $\gamma_{\rm lw} = 1.0$ whereas the dashed lines correspond to the case $\gamma_{\rm lw} = 2.0$. The upper limits from *Fermi*-LAT 2LAC and 3LAC analysis are illustrated as magenta lines and green lines, respectively. In this figure, we showed also the all-flavor neutrinos flux (red points, [32, 189]), the 6-year

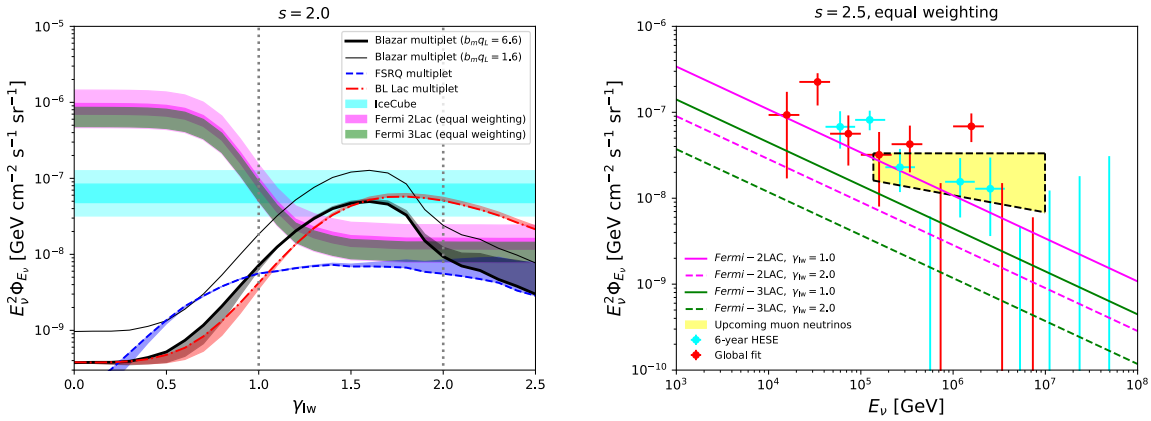

Figure 7.2. All curves and data points in this figure illustrate all-flavor neutrino fluxes. Left panel: Stacking constraints on the contributions of all blazars to the cumulative neutrino flux ($L_{\rm ph,min} = 10^{42}$ erg s^{-1} is used) and high-energy neutrino multiplet constraints on the blazar contributions in the neutrino sky for an ε_ν^{-2} neutrino spectrum. The magenta and green areas correspond to the all-blazar upper limit from *Fermi*-LAT-2LAC and *Fermi*−3LAC equal weighting analysis, respectively. The cyan horizontal area shows the cumulative neutrino flux detected by IceCube. The blue dashed, red dash-dotted and thick black lines illustrate the $m \geq 2$ multiplet constraints for FSRQs, BL Lacs and all blazars whereas the corresponding areas show the uncertainties. The thin black line is the $m \geq 3$ multiplet constraint for all blazars. Right panel: the energy-dependent upper limits from the stacking analysis for the all-blazar contributions, assuming a neutrino spectral index $s = 2.5$.

high-energy starting events (cyan points, [190]) and the the best fit to the upcoming muon neutrinos scaled to three-flavor case (yellow area). The previous discussion reveals that $\mathcal{F}(\gamma_{\rm lw})$ may depend on $L_{\rm ph,min}$ moderately, when $\gamma_{\rm lw}$ is smaller than 1.0. We will further demonstrate in Sec. 7.3 that, in the range of $\gamma_{\rm lw} \lesssim 1.0$, the neutrino multiplet constraints are more stringent than the upper limits derived from the stacking analyses, which manifests its complementarity in constraining the cumulative neutrino flux from all blazars over a wide range of $\gamma_{\rm lw}$.

7.3 Implications of High-Energy Neutrino Multiplet Limits

Here, we present another type of constraints on the origins of IceCube diffuse neutrinos, using the negative results from the clustering test of neutrino-induced muon track events. These high-energy track events are generally detected by IceCube with the angular resolution ~ 0.5 deg, which enables us to determine the incoming directions and perform

clustering analysis on their time and spatial distributions. So far, all the clustering tests based on high-energy muon neutrinos have found no statistically significant evidence of clustering in the arrival distribution of neutrinos [266, 467–469].

In this section, we investigate the implications of the non-detection of neutrino multiplet sources, and consider the limits on blazar contributions to the cumulative neutrino background. To achieve this goal, we follow [131] and write down the limits on the effective source densities. The formalism presented by [131] is applicable to blazars with a general luminosity weighting $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$ since the functions $L_\nu(dN_{\rm bl}/dL_{\rm ph}) \propto (L_{\rm ph})^{\gamma_{\rm lw}+1}\phi_{\rm bl}$ are sharply peaked around some effective luminosities $L_{\rm ph}^{\rm eff}$, which demonstrates that the effective source densities and the neutrino luminosity densities are well defined and constrained. Below, we define these crucial quantities and derive the neutrino multiplet constraints for our blazar case.

Assuming the number of sources that produce more than $k-1$ multiplet events is $N_{m\geq k}$, the constraint from the non-detection of $m \geq k$ multiplet events can be obtained by requiring $N_{m\geq k} \leq 1$. [131] studied the implications to the neutrino sources using the absence of "high-energy" multiplet neutrino sources, and calculated the upper limit on the local source number density for an ε_ν^{-2} neutrino spectrum,

$$
\begin{aligned}
n_0^{\rm eff} \lesssim 1.9\times10^{-10}\ {\rm Mpc}^{-3} &\left(\frac{\varepsilon_\nu L_{\varepsilon_\nu}^{\rm ave}}{10^{44}\ {\rm erg\ s^{-1}}}\right)^{-3/2}\left(\frac{b_m q_L}{6.6}\right)^{-1} \\
&\times\left(\frac{F_{\rm lim}}{10^{-9.2}\ {\rm GeV\ cm^{-2}\ s^{-1}}}\right)^{3/2}\left(\frac{2\pi}{\Delta\Omega}\right),
\end{aligned}
\tag{7.9}
$$

where $\varepsilon_\nu L_{\varepsilon_\nu}^{\rm ave}$ is the time-averaged neutrino luminosity of the source, $F_{\rm lim} \sim (5-6)\times 10^{-10}$ GeV cm^{-2} s^{-1} is the 8-year IceCube point-source sensitivity at the 90% confidence level [470], $q_L \sim 1-3$ denotes a luminosity-dependent correction factor, $\Delta\Omega$ represents the sky coverage of the detector and the details of $m \geq k$ neutrino multiplet constraints are encoded in the factor b_m. [131] find $b_m \simeq 6.6$ for $m \geq 2$ multiplets and $b_m \simeq 1.6$ for triplets or higher multiplets (e.g., $m \geq 3$). Note that the point-source sensitivity enters the above expression but the numerical results are obtained by calculating the number of tracks using the muon effective area [131].

The purpose of this work is to explore the implications for blazar models using existing equations from previous work without making new analyses on multiplet sources. We simply use the results of the previous analysis by [131], which gives the upper limit on the effective number density, $n_0^{\rm eff}(\varepsilon_\nu L_{\varepsilon_\nu}^{\rm ave})$. Moreover, another reason that we choose this approach is that these results are also consistent with the latest limits on transient

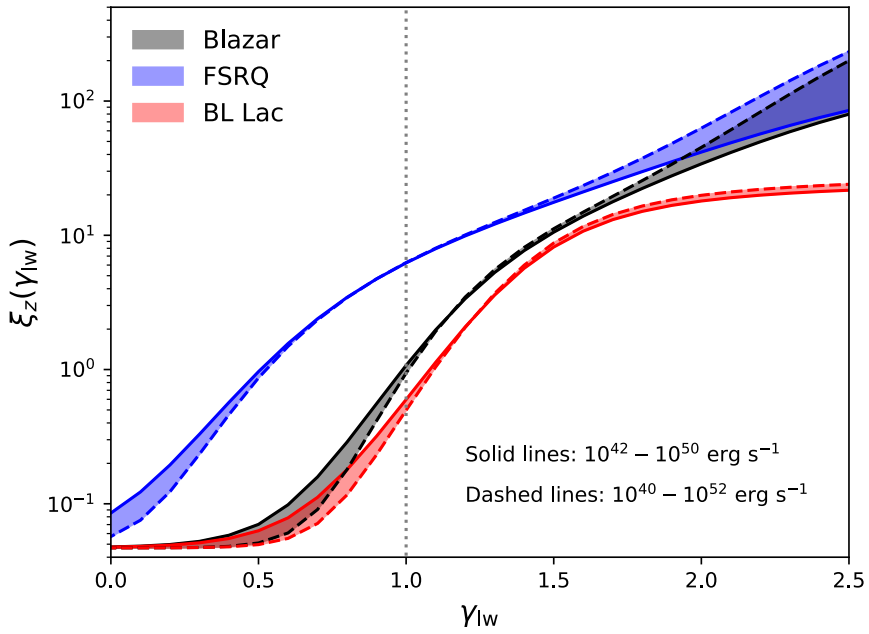

Figure 7.3. The redshift evolution factor ξ_z for FSRQs (blue area), BL Lacs (red area) and all blazars (black area). The solid and dashed boundaries correspond to different schemes of $L_{\rm ph,min}$ and $L_{\rm ph,max}$.

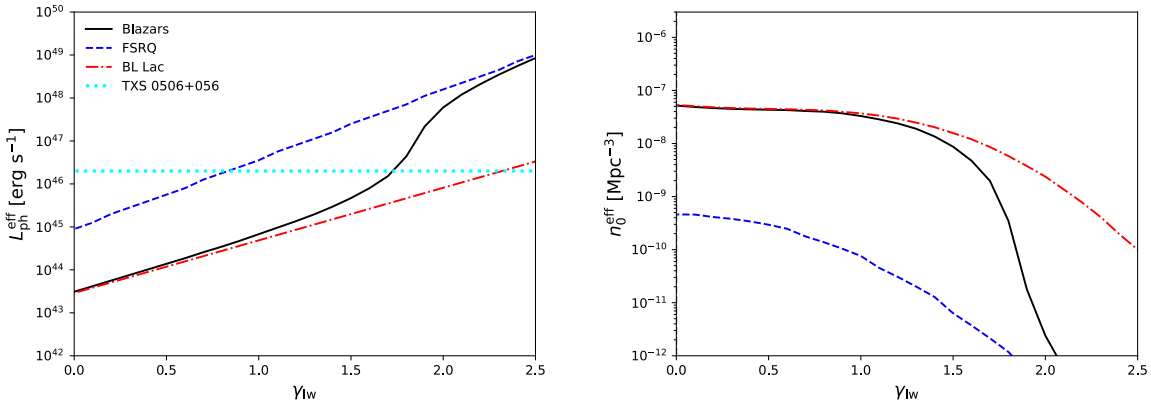

Figure 7.4. Left panel: The effective gamma-ray luminosity for FSRQs (blue dashed line), BL Lacs (red dash-dotted line) and all blazars (black line). The dotted horizontal line indicates the luminosity of TXS 0506+056, one blazar that features an intermediate luminosity, $L_{\rm TXS} \simeq 10^{46.3}$ erg s^{-1} [8]. Right panel: The effective local number densities for different source classes. The line styles in this panel have the same meaning as the left panel.

sources (after the number density is converted into the rate density, [471]) [1]

One can write the limit on the cumulative neutrino flux from the sources as a function of $n_0^{\rm eff}$ and the redshift evolution factor ξ_z [8]:

$$\begin{aligned} E_\nu^2 \Phi_{E_\nu}^{\rm (m)} &\approx \frac{3\xi_z c t_H}{4\pi} n_0^{\rm eff} (\varepsilon_\nu L_{\varepsilon_\nu}^{\rm ave}) \\ &\lesssim 6.9\times10^{-9}\ {\rm GeV cm^{-2} s^{-1} sr^{-1}} \left(\frac{\Delta\Omega}{2\pi}\right)^{2/3} \left(\frac{\xi_z}{0.7}\right) \\ &\times \left(\frac{b_m q_L}{6.6}\right)^{-2/3} \left(\frac{n_0^{\rm eff}}{10^{-7}\ {\rm Mpc^{-3}}}\right)^{1/3} \\ &\times \left(\frac{F_{\rm lim}}{10^{-9.2}\ {\rm GeV\ cm^{-2}\ s^{-1}}}\right), \end{aligned} \tag{7.10}$$

where t_H is the Hubble time. In this expression, ξ_z represents the redshift weighting of the neutrino luminosity of the sources and can be evaluated through [347]

$$\xi_z(\gamma_{\rm lw}) = \frac{\int dz (1+z)^{-1} \left|\frac{dt}{dz}\right| f(z,\gamma_{\rm lw})}{\int dz \left|\frac{dt}{dz}\right|}, \tag{7.11}$$

where $f(z,\gamma_{\rm lw})$ is the redshift evolution function of the neutrino luminosity density normalized to unity at $z=0$ for the luminosity correlation $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$, e.g., for blazars we have $f^{\rm (bl)}(z,\gamma_{\rm lw}) = [\varepsilon_\nu Q_{\varepsilon_\nu}^{\rm (bl,all)}(z,\gamma_{\rm lw})]/[\varepsilon_\nu Q_{\varepsilon_\nu}^{\rm (bl,all)}(0,\gamma_{\rm lw})]$. Similarly, we can also calculate the ξ_z for the blazar subclasses, FSRQs and BL Lacs using the luminosity functions from [463, 464]. The black, blue and red areas in Fig. 7.3 illustrate the redshift evolution factor $\xi_z(\gamma_{\rm lw})$ for all blazars, FSRQs and BL Lacs, respectively. When $\gamma_{\rm lw}=1$, we find $\xi_z \sim 7-8$ for the gamma-ray luminosity density evolution of FSRQs and $\xi_z \sim 0.6-0.7$ for that of BL Lacs, which are consistent with the values found by [54] and [131]. The solid and dashed boundaries in Fig. 7.3 correspond to the sample schemes, ($L_{\rm ph,min}=10^{42}$ erg s^{-1}, $L_{\rm ph,max}=10^{50}$ erg s^{-1}) and ($L_{\rm ph,min}=10^{40}$ erg s^{-1}, $L_{\rm ph,max}=10^{52}$ erg s^{-1}), respectively. If $\gamma_{\rm lw}$ is lower than 1.0, low-luminosity sources at lower redshift contribute a significant component to the total neutrino luminosity density, therefore, a smaller $L_{\rm ph,min}$ results in a smaller ξ_z. On the contrary, a strong luminosity correlation

[1] The limit on the rate density of neutrino transients accounting for the diffuse flux is $\rho_0^{\rm eff} \gtrsim 1.7\times10^4\,{\rm Gpc^{-3}\,yr^{-1}}\,(b_m q_L/6.6)^2(\Delta\Omega/2\pi)^2(T_{\rm obs}/8\ {\rm yr})^2(\xi_z/0.7)^{-3}\phi_{\rm lim,-1}^{-3}$ $\max[N_{\rm fl},1]$, where $N_{\rm fl}\approx f_{\rm fl}T_{\rm obs}/t_{\rm dur}\approx T_{\rm obs}/\Delta T_{\rm fl}$ is the number of flaring periods and $\Delta T_{\rm fl}$ is the typical flare interval [8]. For $\Delta T_{\rm fl}\lesssim T_{\rm obs}$, the density and diffuse limits become similar to those for steady sources. Substituting the time-averaged sensitivity gives conservative results because of $F_{\rm lim}>\phi_{\rm lim}/T_{\rm obs}$. For $\Delta T_{\rm fl}\gtrsim T_{\rm obs}$, we expect $\rho_0^{\rm eff}T_{\rm obs}\approx n_0^{\rm eff}(T_{\rm obs}/\Delta T_{\rm fl})$. Because of $n_0^{\rm eff}(T_{\rm obs}/\Delta T_{\rm fl})\lesssim n_0^{\rm eff}$, the limits for steady sources can be regarded as conservative.

with $\gamma_{\rm lw} \gtrsim 1.5$ boosts the contribution from high-redshift bright blazars, which leads to a larger $f(z, \gamma_{\rm lw})$ at higher redshift and as a result makes ξ_z larger, as $L_{\rm ph,max}$ increases.

Besides the factor ξ_z, it is also necessary to calculate the effective local number density $n_0^{\rm eff}$, which characterizes the the number density of sources that dominate the neutrino luminosity density for each specified source population. In this work, we use the luminosity functions in combination with the luminosity weighting relation $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$ to estimate the effective number densities $n_0^{\rm eff}$ for blazars, FSRQs and BL Lacs. Here, we follow the procedure presented by [131]. For each class of neutrino sources, we define an effective neutrino luminosity $L_\nu^{\rm eff} \propto (L_{\rm ph}^{\rm eff})^{\gamma_{\rm lw}}$ using the corresponding effective gamma-ray luminosity $L_{\rm ph}^{\rm eff}$ obtained by maximizing $(L_{\rm ph})^{\gamma_{\rm lw}}(dN/d\ln L_{\rm ph}) = (L_{\rm ph})^{\gamma_{\rm lw}+1}\phi(L_{\rm ph}, z=0)$, where $\phi(L_{\rm ph}, z=0)$ is the local luminosity function of the sources that we are interested in. Since the function $(L_{\rm ph})^{\gamma_{\rm lw}+1}\phi(L_{\rm ph}, z=0)$ has a maximum around its extreme point for each source population, we may regard blazars, FSRQs and BL Lacs as "quasi-standard candle" sources, among which the neutrino productions are dominated by the sources distributed closely around one certain effective luminosity $L_{\rm ph}^{\rm eff}$. In this case, we have justified the applicability of the equation appeared in this section to constrain the neutrino fluxes from blazars and the subclasses. The left panel of Fig. 7.4 shows the effective gamma-ray luminosity densities for all blazars (black solid line), FSRQs (blue dashed line) and BL Lacs (red dash-dotted line). Intuitively, $L_{\rm ph}^{\rm eff}$ of FSRQ should be larger than that of BL Lacs since FSRQs are more luminous than BL Lacs. Moreover, the function $(L_{\rm ph})^{\gamma_{\rm lw}+1}\phi(L_{\rm ph}, z=0)$ achieves its maximum at higher luminosity as $\gamma_{\rm lw}$ increases, which naturally explains the monotonic increase of $L_{\rm ph}^{\rm eff}(\gamma_{\rm lw})$. Considering that low-luminosity BL Lacs dominate the neutrino luminosity density if the luminosity correlation is weak (e.g., $\gamma_{\rm lw} \lesssim 1$) whereas bright FSRQs become increasingly important as $\gamma_{\rm lw}$ increases, the blazar effective luminosity $L_{\rm ph}^{\rm eff}$ converges to the BL Lac case when $\gamma_{\rm lw}$ is less than 1.0 and then gradually approaches to the FSRQ curve, as is confirmed in Fig. 7.4. With the effective neutrino/gamma-ray luminosity, we can write down the effective local number density of the sources

$$n_0^{\rm eff} = \frac{1}{L_\nu^{\rm eff}} \int dL_{\rm ph} L_\nu(L_{\rm ph}) \phi(L_{\rm ph}, 0). \tag{7.12}$$

The right panel of Fig. 7.4 shows the effective number densities of all blazars (black solid line), FSRQs (blue dashed line) and BL Lacs (red dash-dotted line). As expected, BL Lacs dominate the number density and the blazar effective number density converges to BL Lac and FSRQ curves respectively when $\gamma_{\rm lw} \lesssim 1.0$ and $\gamma_{\rm lw} \gtrsim 2.0$. Different from

$\mathcal{F}(\gamma_{\rm lw})$ and ξ_z, $L_{\rm ph}^{\rm eff}$ and $n_0^{\rm eff}$ does not depend sensitively on the value of $L_{\rm ph,min}$ and $L_{\rm ph,max}$ in the range $0 \lesssim \gamma_{\rm lw} \lesssim 2.5$. To interpret this, we need to keep in mind that the former two quantities are determined by the integrations over $L_{\rm ph}$, while $L_{\rm ph}^{\rm eff}$ depends only on the shape/slope of the function $(L_{\rm ph})^{\gamma_{\rm lw}+1}\phi(L_{\rm ph}, z=0)$. From the left panel of Fig. 7.4, we find that $L_{\rm ph}^{\rm eff}$ lies roughly in the range $10^{43} - 10^{49}$ erg s^{-1} which is covered by the interval $10^{42} - 10^{50}$ erg s^{-1}, the fiducial range used in our calculation. Meanwhile, the integrand in equation 7.12 peaks around $L_{\rm ph}^{\rm eff}$, therefore once the peak is included, the effective number density $n_0^{\rm eff}$ will not vary too much as the lower and upper bounds of the integral changes.

The above calculations provide the preliminary work and the ingredients needed for calculating the neutrino multiplet limits. Selecting $b_m q_L \simeq 6.6$ for $m \geq 2$ multiplets and $F_{\rm lim} \simeq 10^{9.2}$ GeV cm^{-2} s^{-1} for an ε_ν^{-2} neutrino spectrum, the blue dashed, red dashed-dotted and thick black lines in the left panel of Fig. 7.2 illustrate the neutrino multiplet limits for FSRQs, BL Lacs and all blazars, respectively. The blue, red and black areas shows the corresponding uncertainties due to $L_{\rm ph,min}$ and $L_{\rm ph,max}$, as discussed before. From this figure we find that the all-blazar multiplet constraint converges to the FSRQ case at higher $\gamma_{\rm lw}$ and to the BL Lac case if $\gamma_{\rm lw}$ is less than 1.0, just as expected. We also considered the upper limits for triplet or higher multiplets ($m \geq 3$) by changing the value of $b_m q_L$ to 1.6. In this case, the constraints relax to the thin black line. This consequence can be interpreted as the concession of allowing blazars to produce $m = 2$ multiplet events. So far, all calculations on the multiplet constraints were based on the ε_ν^{-2} neutrino spectrum, and to extend the results to a general spectrum, e.g., $s = 2.5$, detailed calculations on $F_{\rm lim}$ and $n_0^{\rm eff}$ (equation 7.12) are needed, and our results are conservative in this point. Therefore, in the right panel of Fig. 7.2, only upper limits inferred from stacking analysis are shown.

7.4 Discussion

In this paper, we considered how two types of analyses, namely stacking and multiplets, constrain on the contribution of blazars to the cumulative neutrino flux, assuming a generalized luminosity weighting $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$. Using the gamma-ray luminosity functions for blazars, FSRQs and BL Lacs, we estimated the ratio of the neutrino fluxes from *Fermi*-LAT-resolved blazars and from all blazars (including unresolved ones), $\mathcal{F}(\gamma_{\rm lw})$, and the effective number densities, $n_0^{\rm eff}(\gamma_{\rm lw})$, and the redshift evolution factor, ξ_z, for different source classes. The joint use of a stacking and multiplet analysis,

as well as the use of a generalized luminosity function and inclusion of the effect of unresolved blazars, are new aspects which distinguish this analysis from previous ones. The main results are summarized in Figure 7.2. From this figure we found that the multiplet constraints are the most important at lower values of $\gamma_{\rm lw}$, e.g. $\gamma_{\rm lw} \lesssim 1.0$, whereas all-blazar constraints derived from the existing stacking upper limits are more stringent for a stronger luminosity correlation, e.g., $\gamma_{\rm lw} \gtrsim 1.5$. The joint consideration of these two kinds of limits supports the extended argument that all blazars, including $Fermi-$unresolved ones, are unlikely to dominate the cumulative neutrino background for a generic correlation between the neutrino and gamma-ray luminosities, $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$, with the index $0 \lesssim \gamma_{\rm lw} \lesssim 2.5$. Canonical blazar models, which are physically motivated and based on the leptonic scenario, predict $\gamma_{\rm lw} \sim 1.5 - 2.0$ [54]. Our results suggest that the stacking constraints are the most stringent for such physically motivated cases. The multiplet and stacking limits are "complementary", in the sense that these methods have their own advantages in different regimes, and in combination they provide a stronger and tighter constraint than previously, over a wide range of $\gamma_{\rm lw}$, as pointed out by [8]. We also found that while the multiplet constraints are weaker at larger values of $\gamma_{\rm lw}$ they become more stringent again for $\gamma_{\rm lw} \gtrsim 1.5$ due to the rapid decrease of the effective source density.

In this work, we focus on power-law spectra. The limits are stringent for the neutrino flux in the 0.1 PeV range and become weaker at higher energies. For example, neutrino multiplet limits are weaker if one is interested in the origin of ~ 1 PeV neutrinos [8,130,131]. It is possible for blazars to explain the dominant fraction of PeV neutrinos by introducing a lower-energy cutoff of the proton maximum energy [458], although neutrinos at 0.1 PeV and lower energies should come from another population of the sources [472].

One of the uncertainties in this work comes primarily from the selection of the lower and upper limits of the luminosity integral, $L_{\rm ph,min}$ and $L_{\rm ph,max}$. As discussed above, we showed that these uncertainties are well controlled, and the final results are reliable if $L_{\rm ph,min}$ and $L_{\rm ph,max}$ are selected in the fiducial ranges $10^{40} - 10^{42}$ erg s^{-1} and $10^{50} - 10^{52}$ erg s^{-1}, respectively. From the joint constraints illustrated in Fig. 7.2, we conclude that blazars are disfavored as a dominant source of the cumulative neutrino flux measured by IceCube for a luminosity weighting $L_\nu \propto (L_{\rm ph})^{\gamma_{\rm lw}}$ with $0.0 \lesssim \gamma_{\rm lw} \lesssim 2.5$. Since different blazar models considered for explaining the cumulative neutrino flux can be commonly characterized by the correlation index $\gamma_{\rm lw}$ within this range, our calculations on the upper limits and effective number densities would provide rather general constraints for future studies of blazar neutrinos.

Acknowledgements

We thank Marco Ajello for useful discussion on the usage of the luminosity function and Nick Rodd for the useful communication. The work of K.M. is supported by the Alfred P. Sloan Foundation and NSF grants No. PHY-1620777 and No. AST-1908689, while that of C.C.Y and P.M. is supported by the Eberly Foundation.

Disclaimer

The findings and conclusions do not necessarily reflect the view of the funding agencies.