Late-Forming PBH: Beyond the CMB era

Philip Lu,1,* Kiyoharu Kawana,2,† and Alexander Kusenko3,4
1 Center for Theoretical Physics, Department of Physics and Astronomy, Seoul National University, Seoul 08826, Korea
2 School of Physics, KIAS, Seoul 02455, Korea
3 Department of Physics and Astronomy, University of California, Los Angeles Los Angeles, California, 90095-1547, USA
4 Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

The intermediate mass black hole range, $10^{-3} \lesssim M_{\text{PBH}}/M_\odot \lesssim 10^5$, has long offered enticing possibilities for primordial black holes (PBH), with populations in this range postulated to be responsible for some of the black hole binary merger detected events as well as the existence of supermassive black holes embedded at galactic centers. However, a prominent bound derived from PBH accretion during recombination severely restricts the mass fraction of intermediate mass PBH. We address this problem by proposing a formation scenario in which “primordial” black holes form late in our cosmological history, beyond the CMB era, and bypassing this bound. During this crucial epoch, our population of compact objects exist as thermal balls supported by thermal pressure, which eventually cool to Fermi balls supported by degeneracy pressure and finally collapse to PBH. Our mechanism is a viable production method for both the mass gap LIGO-VIRGO-KAGRA detections and the JWST observation of an early time $z > 10$ supermassive black hole. Furthermore, we present the remarkable possibility of PBH formation after the present era, which we term future PBH. Such a population would evade most, if not all bounds on the PBH mass spectrum in the literature and open up previously unthought-of possibilities. Light future PBH could form below the Hawking evaporation threshold and convert the bulk of the matter in the Universe into radiation.

I. INTRODUCTION

The intermediate mass black hole (IMBH) range $\sim 10^{-3} - 10^5 M_\odot$ has long been considered in connection with primordial black hole (PBH) formation, offering many unique possibilities and consequences. Although PBH as the dominant dark matter component in this mass range is ruled out by multiple unique bounds [1–17], a subdominant population can contribute to the LIGO-Virgo-Kagra (LVK) spectrum of detected binary mergers [18–31], act as seeds for the supermassive black holes (SMBH) found at the center of galaxies [32–34], and generate hypervelocity stars through three-body scattering at our galactic center [35–37].

The strongest constraint in this region comes from analysis of accreting PBH, which affect recombination and distort the CMB spectrum [38]. The very restrictive limits of the CMB bound on IMBH pose problems for the PBH interpretation of PBH merger events first detected by LIGO ($f_{\text{PBH}} \sim 10^{-3} - 10^{-4}$ and $M \sim 10-100 M_\odot$) [24] and as supermassive seeds ($f_{\text{PBH}} \lesssim 10^{-5}$ and $M \sim 10^4-10^9 M_\odot$) [32–34]. However, this bound does not apply to “primordial” black holes formed after the CMB1.

We propose a mechanism for producing these late-forming PBHs which circumvent the CMB bound. Using the asymmetric dark fermion model developed in Ref. [40–44], we consider first order phase transitions (FOPT) that produce intermediate mass compact objects existing as thermal balls [44] during recombination. These thermal balls steadily cool, eventually collapsing into Fermi balls then PBH by the present day. Although such a PBH population is prevented from being the dominant DM component by bounds from microlensing, dwarf galaxy heating, etc., the parameter space for PBH LVK binary progenitors and supermassive seeds is considerably freed. Of particular interest are the 70-80 $M_\odot$ merger event [45] and the JWST observation of an active galactic nuclei at $z \approx 10.6$ [46].

In this paper, we develop a mechanism for producing late-forming PBH. In section II A, we outline the basic model and thermal history of compact objects which transform into intermediate mass PBH after recombination. In section III A, the conditions for the transition from thermal remnant to Fermi ball and finally to PBH are derived. In section III B, cooling rates and transition temperatures are calculated. In section IV, the relevant contraints on late-forming PBH are discussed. We summarize in Section V.

II. OUTLINE

We summarize the thermal history of PBH formation in our model and present our main points. The essential point in the present PBH scenario is the existence of a FOPT in the dark sector at low temperatures $T_{\text{SM}} \sim 1 \text{ keV}$, which results in the formation of remnants and its collapse to PBHs. After decoupling from the SM plasma around the EW scale, the dark sector temperature $T_D$ evolves independently of the SM
Decoupling of dark sector inflationary reheating. 2

thermal equilibrium with the same temperature \( T \) following thermal history of the Universe: We first specify a concrete shape of \( U_\chi \) is schematically summarized in Fig. 1, and detailed history here to avoid confusion. The thermal history temperature \( T_{\text{SM}} \), and we outline the whole thermal history here to avoid confusion. The thermal history is schematically summarized in Fig. 1, and detailed calculations are presented in Section III.

A. Thermal History

We use the model of Ref. [40–44] which contains a nearly dark sector with fermions \( \chi, \bar{\chi} \) and a scalar \( \phi \) which couples to the Higgs,

\[
\mathcal{L} = \mathcal{L}_{\text{SM}} - \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{\mu^2}{2} \phi^2 - \frac{\kappa}{2} \phi^2 (H^\dagger H) - U(\phi)
\]

where the potential \( U(\phi) \) is assumed to trigger the FOPT. We represent its thermal effective potential as \( \Delta \rho(\phi, T) \), induces a large mass gap \( \delta m_\phi, \delta m_\chi \gg T_{D} \) between the true and false vacuum, and dark sector particles are trapped in the false vacuum pockets of the expanding true vacuum bubbles with the trapping fraction \( F_\chi^{\text{trap}} \approx 1 \). The strength parameter \( \alpha_D \) of the FOPT is given by

\[
\alpha_D := \frac{1}{\rho_D(T_D)} \left[ \Delta U_{\text{eff}}(T_D) + T_D \frac{\partial \Delta U_{\text{eff}}(T_D)}{\partial T} \right]_{T_D=T_{D^*}}^T,
\]

where \( \rho_D(T) \) is the energy density of dark sector and \( \Delta U_{\text{eff}}(T_D) = U_{\text{eff}}(0, T_D) - U_{\text{eff}}(\phi(T_D), T_D) \). The other strength parameter \( \beta \) which determines the duration of FOPT is given by

\[
\frac{\beta}{H_*} = -T_D \left. \frac{\partial T_D}{\partial T} \right|_{T_D=T_{D^*}},
\]

where \( H_* = H|_{T_{\text{SM}} = T_{\text{SM}},*} \) and \( \Gamma(T_D) \) is the decay rate of the false vacuum. In this paper, we typically consider \( 0.1 \lesssim \alpha \lesssim 1 \) and \( \beta/H_* \sim 100 \), where \( \alpha \) is similarly defined as in Eq. (4) but relative to the SM energy density.

2. The separated false-vacuum pockets at the remnant percolation time, when the false vacuum fraction \( f_\chi \approx 0.29 \), constitute individual remnants with their typical size [42]

\[
R_* \approx \frac{v_w}{\beta}.
\]

where \( v_w \) is the bubble wall velocity. The corresponding asymmetrical number density of \( \chi/\bar{\chi} \) fermions trapped within the remnants is directly related to the eventual mass of the PBH (see Section II.B). The latent heat
reduced increases the temperature of the dark sector to $T_{D,0}$.

3. The remnants undergo an initial shrinking phase [44] driven by the vacuum pressure, releasing latent heat and increasing the temperature of the dark sector in the false vacuum until the outward thermal pressure balances the inward vacuum pressure, forming thermal balls.\(^3\) Since we consider a rapid phase transition with transition parameter $\beta/H_* > 1$, the SM temperature stays roughly constant at $T_{SM} = T_{SM*}$, whereas the dark sector temperature at pressure balance increases to [44]

$$T_{D,1} = \left(\frac{90 \Delta U_{\text{eff}}}{\pi^2 g_D} \right)^{1/4}.$$  \hspace{1cm} (7)

We denote the corresponding terminal radius of the remnant in the initial shrinking as $R_1$, which is given by [44]

$$R_1 \approx \left(\frac{1 + \alpha_D}{4 \alpha_D}\right)^{1/3} R_\ast.$$  \hspace{1cm} (8)

4. The thermal balls are unable to directly collapse to PBH, but remain at a constant temperature $T_{D,1}$ throughout a second slow cooling phase. As they cool, the remnants shrink but are kept in homeostasis at $T_{D,1}$ by the continuous release of latent heat.

5. Eventually, the asymmetrical population of dark fermions, which remain constant in the thermal balls, dominate the thermal population which decreases in proportion to the remnant volume. The thermal ball then makes a transition to a Fermi ball, or a remnant supported by Fermi degeneracy pressure [40].

6. After minimal further cooling, the range of the $\phi$-mediated Yukawa force

$$L_\phi(T_D) = m_\phi^{-1}(T_D) = \frac{1}{\sqrt{\mu^2 + cT_D^2}}$$  \hspace{1cm} (9)

becomes long enough to cause an instability in the Fermi ball, which rapidly transitions to a PBH [41]. In our scenario, this final transition can happen after recombination $T_{SM} \sim 0.3$ eV.

\(^3\) In Ref. [44], an extremely small value of $\kappa$ was assumed to forbid the rapid annihilation $\phi \phi \rightarrow HH$. In the present case, $\kappa$ is still limited to small values $\ll 10^{-16}$ to avoid an extra contribution to the scalar mass after the EW symmetry breaking.

---

**FIG. 2.** Predicted PBH mass (blue) and abundance (orange). Here we take $\eta_\chi = 10^{-7}$, $v_w = 0.6$, $\alpha_D = 0.1$.

### B. PBH Mass and Abundance

The average mass of the black hole resulting from Fermi ball collapse is [41]

$$M_{PBH} \approx 10^2 M_\odot \times \alpha D^{1/4} v_w^3 F_{\chi}^{\text{trap}} \left(\frac{\eta_\chi}{10^{-5}}\right) \left(\frac{\beta/H_*}{100}\right)^{-3} \times \left(\frac{g_D}{4}\right)^{-1/4} \left(\frac{g_{SM*}}{g_{SM,\text{dec}}}\right)^{-2/3} \left(\frac{T_{SM*}}{1 \text{ keV}}\right)^{-2}.$$  \hspace{1cm} (10)

The present day PBH fraction of DM, $f_{PBH} := \rho_{PBH}/\rho_{DM}$ is [41]

$$f_{PBH} \approx 0.1 \left(\frac{M}{10^2 M_\odot}\right) v_w^{-3} \left(\frac{g_D}{4}\right)^{1/2} \left(\frac{g_{SM*}}{g_{SM,\text{dec}}}\right) \times \left(\frac{T_{SM*}}{1 \text{ keV}}\right)^{-3} \left(\frac{\beta/H_*}{100}\right)^{-1} \left(\frac{\Omega_{DM}}{0.26}\right)^{-1}.$$  \hspace{1cm} (11)

In Fig. 2, we show the contours of $M_{PBH}$ (blue) and $f_{PBH}$ (orange) in the $\beta/H_*$ vs $T_{SM*}$ plane. Other parameters are chosen as

$$\eta_\chi = 10^{-6}, \quad v_w = 0.6, \quad \alpha_D = 0.1.$$  \hspace{1cm} (12)

There are two populations of interest which can benefit from evading the CMB bound. First, the hypothetical LVK population of PBH which requires black holes of mass $\lesssim 10^2 M_\odot$ and density $f_{PBH} = 10^{-3} - 10^{-4}$. In our FOPT model, false vacuum remnants are only a fraction of the horizon size. Thus, to
form intermediate mass PBH requires a much lower temperature $T_{\text{SM}} \lesssim \text{keV}$ (much larger horizon) when compared to the standard formation mechanism of horizon-scale perturbation collapse. For the proposed LVK PBH population, the typical parameter space is $\beta/H_* = \mathcal{O}(100)$ and $T_{\text{SM},*} = \mathcal{O}(100 \text{ eV})$. Another interesting parameter space is for SMBH seeds, $M_{\text{PBH}} \gtrsim 10^{10} M_\odot$ and $f_{\text{PBH}} = 10^{-5} - 10^{-4}$, with lower $\beta/H_*$ and $T_{\text{SM},*}$. Since the LVK population is formed at higher temperatures, a more prolonged period is necessary to delay PBH formation, whereas relatively efficient cooling can enable the SMBH seed population to form after recombination.

In Table II, we show two benchmark parameter values, one to represent the hypothetical LVK PBH population and the other SMBH seeds. The corresponding PBH mass spectra is plotted in Fig. 3 against relevant constraints using the results in Ref. [42]. In particular, the yellow region corresponds to the CMB bound coming from the accretion disk emission of PBHs [38], but this does not apply in the present PBH formation scenario. See Section IV for details.

### III. DETAILED THERMAL HISTORY

In the following, we will support our proposed formation scenario with detailed calculations of the cooling rate and collapse conditions. Since such late-forming PBH is a new concept, it is important to work out the viability of the model. We show that our most salient point of delayed PBH formation is well justified.

#### A. Fermi ball Transition and PBH Formation

We first detail the process of PBH formation. After sufficient cooling, the thermal ball forms cold compact objects supported by Fermi degeneracy pressure, Fermi balls, which then collapse into black holes. Here we investigate the transition of thermal balls to Fermi balls and their subsequent collapse due to the Yukawa force [41, 44].

Let us consider the collapse condition for thermal balls, which have a lesser density and therefore are less susceptible to collapse. We assume a uniform density thermal ball with

$$Q_{\text{therm}} = \frac{\zeta(3)\tilde{g}_D T_D^3}{\pi^2 R^3},$$

where $Q_{\text{therm}}$ denotes the total number of particles inside a thermal ball, and $\tilde{g}_D = 1 (3/4)$ for each bosonic (fermionic) specie, and does not include other internal d.o.f. such as spin. The attractive Yukawa energy is given by [41]

$$E_\phi = -\frac{q inspected Q_{\text{therm}}}{20\pi \frac{Q_{\text{therm}}}{R^2}} = -\frac{\tilde{g}_D^2 \zeta(3)^2 (3/4) y_\chi^2 R^3 T_D^3 L_\phi}{3\pi^2},$$

and the resulting pressure is

$$P_\phi = \frac{\partial E_\phi}{\partial V} = \frac{\tilde{g}_D^2 \zeta(3)^2}{2\pi^4} y_\chi^2 T_D^3 L_\phi.$$  

The Fermi ball shrinks if this inward Yukawa pressure is greater than the outward thermal pressure $P_{\text{therm}} = \rho D / 3 = (\pi^2/90)g_D T_D^3$. If the scalar mass $m_\phi(T_D)$ is dominated by the thermal $\chi$ term, i.e. $c = y_\chi^2/6$, then

$$P_\phi = \frac{\tilde{g}_D^2 \zeta(3)^2}{2\pi^4} T_D^3 < \frac{\pi^2 g_D T_D^4}{30} = P_{\text{therm}}$$

for $\tilde{g}_D = \mathcal{O}(1)$. We see that Yukawa pressure cannot overcome thermal pressure, meaning that thermal balls are stable until they cool down sufficiently and become Fermi balls.

The transition point from thermal to degeneracy pressure occurs around the time when the asymmetrical population exceeds the thermal population. The former density is given as [41]

$$n_\chi - n_\chi = \frac{R_*}{R_1}^3 \left[ n_\chi - \tilde{n}_\chi \right]_{T_D = T_{D,*}} = \frac{R_*}{R_1} \times F^{\text{trap}} \eta_\chi s(T_{D,*}),$$

where $s(T_D) = 2\pi^2 g_D T_D^3/45$ is the entropy density of dark sector. While the radius of Fermi ball $R$ steadily shrinks during the cooling phase, the temperature remains steady at $T_{D,1}$ (Eq. (7)) due to pressure balance with the false-vacuum energy. The transition point (radius) is defined by the equality

$$n_\chi - n_\chi = F^{\text{trap}} \eta_\chi \frac{2\pi^4}{45\zeta(3)} \left( \frac{g_D}{g_D} \right) \left( \frac{R_*}{R} \right)^3 \left( \frac{T_{D,*}}{T_{D,1}} \right)^3 = 1$$

This leads to

$$R_{\text{tr}} = 0.4 \times (F^{\text{trap}} \eta_\chi)^{1/3} \left( \frac{g_D}{g_D} \right)^{1/3} \left( \frac{T_{D,*}}{T_{D,1}} \right)^{1/3} R_*$$

where we have used Eq. (7). This is of course comparable to the Fermi-ball radius $R_{\text{FB}}$ [40, 41].

After the transition, the Fermi ball can collapse into PBH if the Yukawa force range $L_\phi$ is longer than the average separation size of $\chi (\bar{\chi})$ particles inside the Fermi-ball [41], which would make the remnant energetically unstable. Explicitly, the condition for Fermi-ball collapse
is given as

\[ L_\phi(T_D) = m_\phi(T_D)^{-1} = \frac{1}{\sqrt{\mu^2 + cT_D^2}} \]

\[ > \frac{1}{y_\chi} \left[ \frac{2\pi}{3\sqrt{3}} \left( \frac{2\pi}{3} \right)^{1/6} R_{FB} Q_{FB}^{1/3} \right] y_\chi^{-1} g_D^{-1/4} T_D^{-1} \alpha_D^{-1/4}, \]

(20)

where \( Q_{FB} \) is the total asymmetric charge inside the Fermi ball [41].

In the simple case where \( R^2 \approx 0 \) and the thermal mass is dominated by the \( \chi \) contribution, \( c = g_\chi^2/3 \), the above condition becomes

\[ T_D/T_{D,1} \lesssim 6g_D^{1/2}\alpha_D^{1/2}. \]

Since \( T_D < T_{D,1} \) after the transition from thermal balls, this result implies that Fermi balls immediately collapse into PBHs for \( 0.01 \lesssim \alpha_D \).\(^4\)

\(^4\) After the transition to Fermi ball, the finite density contribution to the thermal mass is dominated by the \( \chi \) contribution. In this case, the thermal mass is approximately given by \( g_D \rho_D \), where \( g_D \) is the number of degrees of freedom for the lightest species.

### B. Cooling of Thermal Balls

In order to ensure that the PBH forms after recombination, \( T_{SM} \lesssim 1 \text{ eV} \), we calculate the SM temperature at which the transition from thermal ball to Fermi ball to PBH occurs. This temperature depends on the rate of cooling, which can proceed via surface cooling with \( dE/dt = -4\pi R^2 \xi \rho_D \) or volumetric cooling with \( (4\pi/3)R^4 \dot{C} \). Here \( \xi \) is an efficiency parameter where \( \xi = g_l/g_{D,D} \) for blackbody radiation, with \( g_l \) the d.o.f. of the emitted (light) particle species, and \( \dot{C} \) is the cooling rate per unit volume. Note that the volumetric cooling occurs only when the interactions among dark-sector particles are very weak (see Appendix A). For comparison with the recombination era, it is convenient to track the evolution of the radius with the SM temperature \( T_{SM} \).

The evolution equation is

\[ \frac{dR}{dT_{SM}} = \frac{dR}{dT} \frac{dT}{dT_{SM}} = \frac{(45)^{3/2} M_{Pl}}{16\pi^{9/2} g_D g_{SM} R^2 T_{D,1} T_{SM}^3} \frac{dE}{dt}, \]

(22)

where we have used

\[ \frac{dE}{dR} = 4\pi R^2 (\rho_D + \Delta U_{eff}) = \frac{16\pi}{3} R^2 \rho_D \]

\[ = \frac{16\pi}{3} R^2 \times \frac{\pi^2 g_D T_{SM}^4}{30} T_{D,1}, \]

(23)

\[ \frac{dT_{SM}}{dt} = -H T_{SM} \simeq -\frac{1}{M_{Pl}} \sqrt{\frac{8\pi^3 g_{SM} T_{SM}^3}{90}}, \]

(24)

and taken \( g_{SM} \gg g_D \) in the last equality.

First, consider the case of surface cooling, \( dR/dT_{SM} \propto 1/T_{SM}^3 \), which has the solution

\[ R(T_{SM}) = R_1 - \frac{a_s}{2} \left( \frac{1}{T_{SM}^2} - \frac{1}{T_{SM}^2} \right), \]

\[ a_s = \frac{(45)^{3/2} \xi L M_{Pl}}{120\pi^{3/2} g_{SM}^{1/2}}, \]

(25)

where we have assumed that \( g_{SM} = O(1) \) is a constant.

On the other hand, in the volumetric cooling case,

\[ \sim y_\chi (n_\chi - n_\xi)^{1/3} \text{ adds to the scalar mass, and becomes dominant over the thermal contribution as the Fermi ball shrinks. However, we have checked that PBH formation is not prevented because the inward Yukawa pressure is still stronger than the outward Fermi degeneracy pressure.} \]

---

TABLE I. Benchmark parameter points for sample LVK merger and SMBH seed PBH populations.

| \( M_{PBH} \) | \( T_{SM,*} \) | \( y_\chi \) | \( \alpha \) | \( \beta/H_0 \) | \( v_w \) | \( f_{\nu PBH} \) |
|---|---|---|---|---|---|---|
| LVK | 30 \( M_\odot \) | 400 eV | \( 10^{-6} \) | 0.1 | 300 | 0.6 | \( 10^{-3} \) |
| SMBH | \( 3 \times 10^4 \) \( M_\odot \) | 40 eV | \( 10^{-6} \) | 0.1 | 150 | 0.6 | \( 10^{-4} \) |
\[ \frac{dR}{dT_{SM}} \propto R/T_{SM}^3, \] 

the solution is

\[ R(T_{SM}) = R_1 \exp \left[ -\frac{a_v}{2} \left( \frac{1}{T_{SM}^2} - \frac{1}{T_{SM*}^2} \right) \right], \]

\[ a_v = \frac{(45)^{3/2} C M_P}{12\pi^{1/2} g_D 9 T_{SM}^4}. \] \hspace{1cm} (26)

In section III A, we obtained the thermal ball to Fermi ball transition radius \( R_{tr} \) as a function of model parameters, i.e. \( \text{Eq. (19)} \). Combining this result with specific cooling models, we will now estimate the corresponding SM temperature of PBH formation.

1. Blackbody Cooling

If the mass differences of the trapped particles between two phases are negligible, \( \Delta m_i \ll T_{D,1} \), they can freely escape and produce blackbody radiation with \( \xi_l = O(0.1 - 1) \). For large couplings \( y_\chi \) between \( \chi \) and \( \phi \), thermal equilibrium in the dark sector is maintained \([44]\), so that continuous emission of \( \phi \) and \( \chi \) is possible. Since \( R_{tr} \ll R_1 \), the transition temperature can be qualitatively estimated by \( R(T_{SM}) \sim 0 \) and it is given by

\[ T_{SM, tr} \approx \left[ 1 + \frac{8v_w}{3\xi_l} \left( \frac{\beta}{H_*} \right)^{-1} \left( 1 + \frac{\alpha_D}{4\alpha_D} \right)^{1/3} \right]^{-1/2} T_{SM*}. \] \hspace{1cm} (27)

In this case, blackbody radiation rapidly cools and shrinks the remnant so that the transition to Fermi ball and PBH happens shortly after the phase transition. Thus, blackbody cooling is not conducive for late-forming PBH.

2. Evaporation Cooling

On the other hand, if the dark fermions \( \chi, \bar{\chi} \) are only partially trapped \( \Delta m_\chi \gtrsim 5T_D \), then the high energy tail of the Fermi-Dirac distribution may slowly escape and cool the remnant, a scenario studied in Ref. [44]. If the \( \phi \)-mediated \( \chi \bar{\chi} \rightarrow \chi \bar{\chi} \) scattering rate is fast enough to continuously replenish the population of high energy fermions, then evaporation will be surface cooling. On the other hand, if the thermalization timescale of the remnant \( \tau_{therm} \) is longer than the crossing timescale \( t_c \equiv R_1/c \), then evaporation will be volumetric cooling. Here we consider the stronger surface cooling case and leave the detailed calculation of \( \tau_{therm} \) and volumetric cooling for Appendix A. Note that evaporation cooling from escaping \( \phi \) particles is also possible and is analogous to the \( \chi, \bar{\chi} \) cooling we consider here.

The surface cooling factor \( \xi_l \) of evaporation is given in Ref. [44] as

\[ \xi_l = \frac{120 g_\chi}{7\pi^4 g_D} \left( \frac{M_\chi}{T_{D,1}} \right)^3 e^{-M_\chi/T_{D,1}} \] \hspace{1cm} (28)

with \( g_\chi = 4 \) when there are only \( \chi \) and \( \bar{\chi} \) particles. This exponentially suppressed slow cooling rate justifies using the low transition temperature and small transition radius limit \( T_{SM, tr} \ll T_{SM*}, R_{tr} \ll R_1 \) of \( \text{Eq. (25)} \),

\[ T_{SM, tr} = \left( \frac{a_v}{2R_1} \right)^{1/2} \approx T_{SM*} v_w^{-1/2} \left( \frac{\beta}{H_*} \right)^{1/2} \left( \frac{g_\chi}{g_D} \right)^{1/2} \times \left( 1 + \frac{\alpha_D}{4\alpha_D} \right)^{-1/6} \left( \frac{M_\chi}{T_{D,1}} \right)^{3/2} e^{-M_\chi/T_{D,1}}. \] \hspace{1cm} (29)

Then to get a transition temperature between the CMB era and the present day from an initial phase transition temperature \( T_{SM*} \sim 1 \) keV, a reduction of \( 10^3 - 10^6 \) requires \( 20 \lesssim M_\chi/T_{D,1} \lesssim 40 \). For even larger mass to temperature ratios, there is the interesting possibility of producing PBH from collapsing thermal balls in the future. As explained in Section IV, thermal balls are difficult to bound due to their less compact nature. Present day thermal balls would therefore have significantly reduced constraints across the entire mass spectrum, opening up much larger mass windows. Future populations of PBH, which evade present day bounds, may be a possible DM candidate.

IV. BOUNDS

All the constraints based on microlensing, accretion, etc. that are independent of cosmological history still apply to PBH forming between CMB and the present day (see Ref. [49] for a review). However, future forming PBH which exist today as slowly cooling thermal balls could evade and ameliorate most of the present day bounds on its mass spectrum.

We discuss three constraints of interest: BBN \( \Delta N_{eff} \), CMB bounds, and GW emission from the FOPT. Some other bounds on PBH, such as those from considerations of large scale structure formation, may be affected although a detailed analysis is beyond the scope of this paper. While the Fermi balls are cooling, they resemble decaying dark matter, but the current bounds do not constrain a fraction of such dark matter below \( \sim 0.1 \) \([50]\).

A. BBN \( N_{eff} \)

In the models we consider, the FOPT occurs around the keV-scale so that the dark sector particles were free-floating during the BBN era. The relevant bound is the BBN \( \Delta N_{eff} = n_D/n_\nu \lesssim 0.40 \) \([51]\) limit on the number of extra relativistic species. Although this limit may seem to be grossly violated because of the 2 dark sector fermions \( \chi, \bar{\chi} \) and scalar \( \phi \). However, the weak coupling between the dark sector and the SM results in freeze-out at high temperatures after the EW transition. When the dark sector freezes out at around
the EW scale, the dark sector temperature $T_D$ at BBN compared to the SM temperature $T_{SM}$ is $T_D/T_{SM} = (10.75/106.75)^{1/3} \approx 0.5$ (see Eq. (3)), which corresponds to $\Delta N_{eff} \approx g_D(T_D/T_{SM})^4 \ll 0.4$ for $g_D = \mathcal{O}(1)$. Thus, the constraint by BBN can be easily evaded.

B. CMB Accretion Bound

Here we reevaluate the Planck/CMB bound on PBH \[38, 52, 53\] for thermal ball remnants. These bounds are based largely on accretion, so the lower density and surface gravity thermal remnants are unable to convert the potential energy of accreted material into radiation as efficiently.

In Ref. [38], the accretion disk is modeled as a thin disk with luminosity $L \sim 0.1 \dot{m}$ for high accretion rates $\dot{m} = \dot{M}/\dot{M}_{Eddington}$ and as an advection dominated accretion flow $L \sim 0.01 \dot{m}^2$ for low accretion rates. In each of these accretion models, the luminosity is roughly proportional to the surface gravitational potential, which is calculated at the innermost stable circular orbit radius for PBH, $GM/3R_* = 1/6$ for Schwarzschild black holes. In contrast, the minimum radius for thermal balls is much larger and consequently, the surface gravity is much lower. For simplicity, we assume a uniformly distributed thermal ball so that $M = (4\pi R^3/3)\rho_D$ and $\rho_D = g_D(\pi^2/30)T_D^4$ so that the ratio of surface potentials between thermal remnants and PBH is

$$\frac{(GM/R)_{\text{therm}}}{(GM/R)_{\text{PBH}}} = \frac{4\pi^3}{45} g_D R^2 T_D^4 .$$

The surface potential of the thermal ball is highest immediately after formation and decreases during the slow shrinking phase. The ratio of the potentials at this time is

$$\left.\frac{(GM/R)_{\text{therm}}}{(GM/R)_{\text{PBH}}} \right|_{R=R_i} \approx 2 \times 10^{-5} \left( \frac{g_D}{g_{SM*}} \right) \nu_w^2 \times \left( \frac{\beta/H_*}{100} \right)^2 \left( \frac{H_*}{10^9} \right)^{-2} \left( \frac{1 + \alpha D}{\alpha D} \right)^2 .$$

For typical parameters, the thermal ball surface potential even at its maximum is much weaker than for PBH of similar mass. Thus, the accretion based CMB bounds should be weaker by the same factor, i.e., $f_{\text{therm}} < 10^9 f_{\text{PBH}} \sim 10^{-3}$ at its peak assuming the most stringent conditions. Since the thermal balls are formed around $\sim 1$ keV, by the recombination era of $\sim 1$ eV, the radius will have decreased further, weakening the bound. Additional second-order effects such as the lower temperature of the thermal ball accretion disk result in proportionally less ionizing X-rays emitted. PBH accretion following recombination may still induce spectral distortions into the CMB. However, for PBH formation at $z \lesssim \mathcal{O}(100)$ the gas density should be sufficiently low. Therefore, the CMB bound on PBHs does not significantly apply to pre-collapse thermal balls and is superseded by the other non-cosmological PBH bounds in the IMBH mass range.

C. Gravitational Wave Bounds

Another phenomenological signature of FOPT is the production of stochastic GWs \[54–56\]. In general, there are three sources of the stochastic GWs produced during a FOPT: bubble collisions, sound waves, and magnetohydrodynamic (MHD) turbulence in the plasma. Although the GW strength from these sources have different parameter dependences, their peak frequencies are qualitatively the same,

$$f_{\text{peak}} = \mathcal{O}(10^{-2} - 10^{-3}\text{mHz}) \times \left( \frac{\beta}{H_*} \right) \left( \frac{T_{SM*}}{100 \text{ GeV}} \right) ,$$

which means that the typical peak frequencies in our PBH scenario are $f_{\text{peak}} = \mathcal{O}(10^{-11} - 10^{-12}\text{Hz})$ for $\beta/H_* = 100$ because $T_{SM*} \lesssim 1$ keV. These frequencies are too low even for pulsar timing array experiments such as the International Pulsar Timing Array (IPTA) \[57–60\] and the Square Kilometre Array (SKA) \[61\] to detect. Note that there also exists the bound by CMB observations \[62\] in the low frequency region, $f \lesssim 10^{-15}$ Hz, which is well below the frequencies in our model. Thus, there are no bounds from GW signals in the present scenario. We explicitly show the GW spectra in Fig. 4 for the parameters of the two benchmark models presented in Table II B, using the the numerical fitting functions in Ref. \[54\].

V. CONCLUSION

In this paper, we have presented a viable method for producing late-forming PBH which exist in thermal ball form until late in our cosmological history. The primary benefit of this novel production scenario is to avoid constraints on the PBH mass spectrums which
rely on cosmology, most notably the CMB bound in the intermediate mass range. The bypassing of this constraint allows otherwise restricted “primordial” populations to generate some of the (higher mass) merger events detected by LVK and to seed the puzzlingly supermassive black holes found at the center of galaxies and clusters. To support our claims, we have presented detailed calculations of collapse conditions and cooling rates to show that such prolonged formation is indeed possible. Building upon previous works in this series, Refs. [41, 42, 44], we have selected sets of model parameters that generate extended mass distributions of both LVK and SMBH seed populations, showing our results in Fig. 3. We provide a valid formation mechanism for the LVK mass gap black holes as well as the JWST $z \approx 10.6$ SMBH that is unencumbered by CMB bounds.

In addition to PBH forming between the CMB era and the present day, we propose an even more exotic possibility of thermal ball remnants collapsing in the future to Fermi balls and PBH. This ghost population of PBH could circumvent an even larger subset of PBH bounds, so that intermediate mass PBH comprising the whole of dark matter could be possible at later epochs. We therefore term these compact objects future PBH. Future PBH could open up many exciting paths in our cosmological evolution. Since light $M < 10^{17}$ g thermal balls would not be affected Hawking evaporation, they would be exempt from limits in that mass range. Future PBH could form around or below the mass threshold of $10^{15}$ g, rapidly emitting Hawking radiation, converting the bulk of the matter density into radiation. We leave exploring the possibility of a second radiation era from future light PBH to future work.

**ACKNOWLEDGEMENTS**

We would like to thank Ke-Pan Xie, TaeHun Kim, and Volodymyr Takhistov, Marco Flores, Edoardo Vitagliano, and Kazunori Kohri for useful discussions. The work of PL is supported by Grant Korea NRF-2019R1C1C1010050. K.K. would like to thank Yukawa Institute for Theoretical Physics, Kyoto University for the support and the hospitality during his stay by the long term visiting program. A.K. was supported by the U.S. Department of Energy (DOE) Grant No. DE-SC0009937, by the Simons Foundation Fellowship, by the World Premier International Research Center Initiative (WPI), MEXT, Japan, by Japan Society for the Promotion of Science (JSPS) KAKENHI grant No. JP20H05853, and by the UC Southern California Hub with funding from the UC National Laboratories division of the University of California Office of the President.

**Appendix A: Volumetric Evaporation Cooling**

Here we calculate the thermalization time $\tau_{\text{therm}}$ which will relate the volumetric cooling rate $\dot{C}$ to the parameters of the FOPT and enable a quantitative condition for the dominance of surface cooling at large couplings (more interactions and faster thermalization) to volumetric cooling at small couplings.

From Ref. [44], the scattering cross section of dark sector particles can be approximated by the dominant process of $\chi \chi \rightarrow \chi \chi$ scattering,

$$\frac{d\sigma}{dt} = \frac{y_{\chi}^4}{16\pi s^{2}}, \quad (A1)$$

where $s = (p_1 + p_2)^2 = 4E_{\text{cm}}^2$ and $t = (p_1 - p_2)^2$ are the usual Mandelstam variables. The energy change $\delta E$ of the colliding particle 1 after its collision with particle 2 is

$$\delta E = \frac{(E_1 - E_2)(\cos \phi - 1)}{2} + \sqrt{E_1 E_2(1 + \cos \theta)/2} \sin \phi, \quad (A2)$$

where $\theta$ is the angle between the particle momenta before the collision in the thermal ball frame and $\phi$ is the (post-)scattering angle in the CM frame. By symmetry, both angles are spherically distributed. We use a thermal Fermi-Dirac distribution for both particles 1 and 2 as a function of $\epsilon = p/T$

$$f(\epsilon) = \frac{1}{e^\epsilon + 1}. \quad (A3)$$

Using the relations

$$\frac{d\sigma}{d\phi} = \frac{d\sigma}{dt} \frac{dt}{d\phi} = \frac{y_{\chi}^4 \sin \phi}{64\pi E_1 E_2 (1 - \cos \theta)}, \quad (A4)$$

and $\epsilon_{\text{rel}} \approx 2\epsilon$, rate of energy transfer of a particle 1 with energy $E_1 = \epsilon_1 T$ is

$$\frac{\delta E_1}{\delta t} = \frac{T^2 y_{\chi}^4}{512\pi^3 \epsilon_1} \int_0^\infty d\epsilon_2 \epsilon_2 e^{\epsilon_2} e^{-\epsilon_1} \left[ \int_0^\pi d\theta \sin \theta \left( \frac{1 - \cos \theta}{1 - \cos \epsilon_1} \right) \right]$$

$$\times \left( \int_0^\pi - \int_0^{2\pi} \right) d\phi \sin^2 \phi \left( \frac{1 - \epsilon_2}{1 + \epsilon_2} \frac{(\epsilon_1 - \epsilon_2)(\cos \phi - 1)}{2} \right)$$

$$+ \sqrt{\epsilon_1 \epsilon_2 (1 + \cos \theta)} \sin \phi = 2.582 \times 10^{-4} \frac{T^2 y_{\chi}^4}{\sqrt{\epsilon_1}}. \quad (A5)$$

To find the thermalization timescale, average across the incident particle energy distribution

$$\tau_{\text{therm}} = \frac{2}{3\zeta(3)} \int_0^\infty d\epsilon_1 \epsilon_1^2 e^{-\epsilon_1} \int_0^\infty d\epsilon_2 E_1 E_2 \frac{1}{\epsilon_1} \frac{1}{\epsilon_2} e^{-(\epsilon_1 + \epsilon_2)/\epsilon_1} \frac{1}{\epsilon_2} e^{-(\epsilon_1 + \epsilon_2)/\epsilon_2}$$

$$\delta E_1 \frac{\delta E_2}{\delta t} = \frac{2.402 \times 10^4}{T y_{\chi}^2}. \quad (A6)$$

Comparing this to the crossing timescale

$$t_c = R_1/c = \frac{1.22v_w}{\beta} \left( \frac{1 + \alpha_D}{4\alpha_D} \right)^{1/3}, \quad (A7)$$

Here $v_w$ is the sound speed.
at temperature \( T = T_{D,1} \) (Eq. (7)), the cooling will be surface dominated if \( t_c > \tau_{\text{therm}} \),
\[
y_X > 9.50 \times 10^{-6} \left( \frac{T_{\text{SM}}}{1 \text{ keV}} \right)^{1/4} \left( \frac{\beta}{\mathcal{H}} \right)^{1/24} \times g_{\text{SM}}^{1/6} g_{\text{SM,dec}}^{1/12} \left( \frac{40 \rho_D}{1 + \alpha_D} \right)^{1/12},
\]
and volume cooling dominated if the opposite inequality is satisfied.

The volumetric cooling rate \( \dot{C} \) from evaporation cooling is [44]
\[
\dot{C} = \frac{g_Y T_{\text{SM}}^3}{2\pi^2 \tau_{\text{therm}}} e^{-M_X/T} = \frac{g_Y T_{\text{SM}}^3}{4.74 \times 10^5} e^{-M_X/T},
\]
where we have substituted the previous result for \( \tau_{\text{therm}} \).

As with the surface cooling case, the volumetric cooling rate is exponentially suppressed. Therefore the use of the small transition temperature limit of Eq. (26) is justified. The transition temperature in this regime is then
\[
T_{\text{SM, tr}} = \left( \frac{a_\nu}{2 \ln(R_1/R_{\text{tr}})} \right)^{1/2} = 1.7 \text{ eV} \left( \frac{y_X}{10^{-6}} \right)^2 \left( \frac{T_1}{\text{keV}} \right)^{1/2} \times \left( \frac{R_1}{R_{\text{tr}}} \right)^{-1/2} \left( \frac{M_X}{T_1} \right)^{3/2} e^{-M_X/2T_1}.
\]

For this boundary value of \( y_X \), transitioning between the CMB and the present era requires \( 10 \lesssim M_X/T_1 \lesssim 25 \).

**Appendix B: SM Portal Cooling**

The thermal ball remnant can cool by SM particle production via the Higgs portal. Since SM particles are not trapped by the bubble wall, they can easily escape. For most of the cooling, \( T_{\text{SM}} \ll T_{D,1} \), as the remnant is heated up during the initial collapse, the SM plasma within the thermal ball is negligible. Thus, the energy in the produced particles freely disperses into the surrounding low temperature plasma.

The dominant production channel at the low temperatures of interest, \( T \lesssim 1 \text{ keV} \), is \( \phi \phi \rightarrow \gamma \gamma \) with cross section
\[
\sigma \simeq \frac{\alpha^2 e^2 (\kappa \nu)^2}{1152 \pi^3 \sin \theta_W M_W^2 (s - M_W^2)^2} \text{ for } s \ll m_1^2,
\]
\[
\simeq 2.717 \times 10^{-9} \frac{\kappa^2 s}{M_W^2},
\]
Integrating over the incident and target particle Fermi-Dirac momentum distribution, and incoming angle \( \theta \), the cooling rate is calculated to be
\[
\dot{C} = \frac{T^3}{2\pi^2} \int d\epsilon_1 \frac{\epsilon_1^2}{\epsilon_1^2 + 1} \frac{T^3}{2\pi^2} \int d\epsilon_2 \frac{\epsilon_2^2}{\epsilon_2^2 + 1} (\epsilon_1 + \epsilon_2) T \times \int_0^\pi d\theta \frac{\sin \theta}{2} v_{\text{rel}} d\epsilon \epsilon_1 d\epsilon_2 = 5.808 \times 10^{-9} \kappa^2 T^9 M_W^2,
\]
where we have used \( v_{\text{rel}} = 2, s = 2\epsilon_1 \epsilon_2 (1 - \cos \theta) \), and the factor \( (\epsilon_1 + \epsilon_2) T \) accounts for the SM to the SM. The cooling rate is minimal, so that the low transition temperature limit of Eq. (26) gives
\[
T_{\text{SM, tr}} \simeq \left( \frac{a_\nu}{2 \ln(R_1/R_{\text{tr}})} \right)^{1/2} = 2.8 \times 10^{-6} \text{ eV} \kappa \times \left( \frac{T_1}{\text{keV}} \right)^{5/2} \left( \frac{R_1}{R_{\text{tr}}} \right)^{-1/2}.
\]

For the transition temperature to be between CMB \( T_{\text{SM}} \sim 1 \text{ eV} \) and the present day \( T_{\text{SM}} \sim 2 \times 10^{-4} \text{ eV} \), the thermal ball temperature should be in the range of \( 6 \text{ keV} \kappa^{-2/5} \lesssim T_{D,1} \lesssim 170 \text{ keV} \kappa^{-2/5} \). SM portal cooling is therefore not conducive to forming intermediate mass PBH by the present day unless there is extreme reheating of the dark sector particles from the FOPT temperature of \( \sim 1 \text{ keV} \) to \( \sim 1 \text{ TeV} \). However, the extreme slow cooling of the thermal ball remnants is ideal for the scenario of future PBH as discussed in Section III B 2.

[1] P. Lu, V. Tikhistov, G. B. Gelmini, K. Hayashi, Y. Inoue, and A. Kusenko, Astrophys. J. Lett. 908, L23 (2021), 2007.02213.
[2] V. Tikhistov, P. Lu, G. B. Gelmini, K. Hayashi, Y. Inoue, and A. Kusenko, JCAP 03, 017 (2022), 2105.06099.
[3] V. Tikhistov, P. Lu, K. Murase, Y. Inoue, and G. B. Gelmini (2021), 2111.08699.
[4] P. N. Wilkinson, D. R. Henstock, I. W. A. Browne, A. G. Polatidis, P. Augusto, A. C. S. Readhead, T. J. Pearson, W. Xu, G. B. Taylor, and R. C. Vermeulen, Phys. Rev. Lett. 86, 584 (2001), astro-ph/0101328.
[5] R. Ibata, C. Nipoti, A. Sollima, M. Bellazzini, S. C. Chapman, and E. Dalessandro, "Mon. Not. Royal Soc." 428, 3648 (2013), 1210.7787.
[6] J. D. Bradford, M. Geha, R. R. Muñoz, F. A. Santana, J. D. Simon, P. Côte, P. B. Stetson, E. Kirby, and S. G. Djorgovski, Astrophys. J. 743, 167 (2011), 1110.0484.
[7] M. Oguri, J. M. Diego, N. Kaiser, P. L. Kelly, and T. Broadhurst, Phys. Rev. D 97, 023518 (2018), 1710.00418.
[8] Y. Inoue and A. Kusenko, JCAP 1710, 034 (2017), 1705.00791.
[9] B. J. Carr and M. Sakellariadou, Astrophys. J. 516, 195 (1999).
[10] R. Murgia, G. Scelfo, M. Viel, and A. Raccanelli, Phys. Rev. Lett. 123, 071102 (2019), 1903.10509.
[11] T. D. Brandt, Astrophys. J. Lett. 824, L31 (2016),
