ANNUAL REVIEWS

Annual Review of Astronomy and Astrophysics

The Assembly of the First Massive Black Holes

Kohei Inayoshi,[1] Eli Visbal,[2] and Zoltán Haiman[3]

[1]Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China; email: inayoshi@pku.edu.cn

[2]Department of Physics and Astronomy and Ritter Astrophysical Research Center, University of Toledo, Toledo, Ohio 43606, USA; email: Elijah.Visbal@utoledo.edu

[3]Department of Astronomy, Columbia University, New York, NY 10027, USA; email: zoltan@astro.columbia.edu

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020. 58:27–97

First published as a Review in Advance on March 30, 2020

The *Annual Review of Astronomy and Astrophysics* is online at astro.annualreviews.org

https://doi.org/10.1146/annurev-astro-120419-014455

Copyright © 2020 by Annual Reviews. All rights reserved

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

cosmology, first galaxies, active galactic nuclei, quasars

Abstract

The existence of $\sim 10^9\,M_\odot$ supermassive black holes (SMBHs) within the first billion years of the Universe has stimulated numerous ideas for the prompt formation and rapid growth of black holes (BHs) in the early Universe. Here, we review ways in which the seeds of massive BHs may have first assembled, how they may have subsequently grown as massive as $\sim 10^9\,M_\odot$, and how multimessenger observations could distinguish between different SMBH assembly scenarios. We conclude the following:

- The ultrarare $\sim 10^9\,M_\odot$ SMBHs represent only the tip of the iceberg. Early BHs likely fill a continuum from the stellar-mass ($\sim 10\,M_\odot$) to the supermassive ($\sim 10^9$) regimes, reflecting a range of initial masses and growth histories.
- Stellar-mass BHs were likely left behind by the first generation of stars at redshifts as high as ~ 30, but their initial growth typically was stunted due to the shallow potential wells of their host galaxies.
- Conditions in some larger, metal-poor galaxies soon became conducive to the rapid formation and growth of massive seed holes, via gas accretion and by mergers in dense stellar clusters.
- BH masses depend on the environment (such as the number and properties of nearby radiation sources and the local baryonic streaming velocity) and on the metal enrichment and assembly history of the host galaxy.

- Distinguishing between assembly mechanisms will be difficult, but a combination of observations by the *Laser Interferometer Space Antenna* (probing massive BH growth via mergers) and by deep multiwavelength electromagnetic observations (probing growth via gas accretion) is particularly promising.

Contents

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

1. INTRODUCTION

1.1. Summary of Observations

Observations of high-redshift quasars ($z \gtrsim 6$) indicate that supermassive black holes (SMBHs) with masses greater than $\sim 10^9\ M_\odot$ formed within the first billion years after the Big Bang. These objects are very rare (having number density of $\sim 1\ \mathrm{Gpc}^{-3}$) and have so far been found in optical/

IR surveys that cover very large portions of the sky. The Sloan Digital Sky Survey (SDSS) was the first to discover high-redshift quasars (Fan et al. 2001, 2003) and was followed by several additional efforts such as the UKIRT (United Kingdom Infrared Telescope) Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007), the Canada-France High-redshift Quasar Survey (CFHQS; Willott et al. 2007), and the Panoramic Survey Telescope & Rapid Response System 1 (Pan-STARRS1; Morganson et al. 2012) survey. These surveys have yielded well over 100 quasars with redshifts $z > 6$, many of which have inferred black hole (BH) masses of $M_\bullet > 10^9$ $M_\odot$. The most massive of these, SDSS J010013.02+280225.8 (Wu et al. 2015), has an estimated mass of 1.2×10^{10} $M_\odot$ at $z = 6.3$. The most distant, ULAS J1342+0928 (Bañados et al. 2018), has a mass of 7.8×10^8 $M_\odot$ at $z = 7.54$. These large surveys have enabled accurate characterization of the bright end of the quasar luminosity function (LF; e.g., Jiang et al. 2016). Recently, a large sample of additional, lower-luminosity quasars have been uncovered in the Subaru High-z Exploration of Low-Luminosity Quasars (SHELLQs) survey, bringing the total number of $z > 6$ quasars to nearly 200 and extending the constraints on the LF to fainter quasars (Matsuoka et al. 2018c). These observations are summarized in **Figure 1** and will be discussed in detail in Section 2 below.

In addition to optical/IR surveys, quasars have been observed across a variety of wavelengths, from X-rays [with the *Chandra X-ray Observatory*, *X-ray Multi-Mirror Mission-Newton*, and the *Swift X-ray Telescope*] to radio (with the Very Large Array, Giant Metrewave Radio Telescope, and Murchison Widefield Array). For example, the Faint Images of the Radio Sky at Twenty centimeters (FIRST) survey (Becker et al. 1995) found many quasars. Overall, observations across all wavelengths indicate that there is little evolution in the physical properties of the brightest quasars or their host galaxies over cosmic time, as inferred from detailed optical/IR, X-ray (Nanni et al. 2017), or radio (Bañados et al. 2015) analyses, implying that the hosts of these objects formed early as well (see Section 2).

Population III (Pop III) stars: stars formed from primordial gas (metal-free or nearly metal-free) and predicted to have significantly higher masses than metal-enriched stars

Minihalos: low-mass dark matter halos ($\sim 10^{5-6}$ $M_\odot$) expected to host the first stars

Eddington-limited accretion: the accretion rate when there is a balance between radiation pressure and the gravitational force on infalling matter

1.2. Timescale Issues

The presence of $>10^9$ $M_\odot$ SMBHs before the Universe was a billion years old represents an intriguing puzzle. How did the first SMBHs grow so large so fast? This question had been raised already at the discovery of quasars at $4 < z < 5$ (Turner 1991), but pushing the redshift limits to $z > 7$ and the correspondingly shorter cosmic time available made it significantly more intriguing (Haiman & Loeb 2001). A naive explanation is that these early SMBHs were seeded by BH remnants of the first Population III (Pop III) stars. Pop III stars are expected to form in $\sim 10^{5-6}$ $M_\odot$ dark matter (DM) minihalos through primordial gas undergoing molecular hydrogen (H_2) cooling. The metal-free primordial gas is significantly warmer (a few 100 K) than star-forming molecular clouds in the ISM of low-z galaxies ($\sim$10 K). The general expectation is that inefficient cooling of the primordial gas leads to inefficient fragmentation, making Pop III stars unusually massive. The initial mass function (IMF) of Pop III stars remains uncertain, but simulations suggest that it is indeed top-heavy, with a mass range of $10 \lesssim M_\star/M_\odot \lesssim 10^3$ (Hirano et al. 2014).

If BH growth is dominated by Eddington-limited accretion, a seed grows exponentially with an e-folding time of $t_{\rm Edd} \approx 50$ Myr, assuming a radiative efficiency of $\epsilon \approx 10\%$. A comparison between the observed quasar activity across all redshifts and the local population of remnant SMBHs (Soltan 1982) implies that most low-z SMBHs assembled the bulk of their mass at $z = 2$–3 at this efficiency (Haehnelt et al. 1998, Yu & Tremaine 2002, Shankar et al. 2004). This efficiency is also similar to the value ($\sim$0.06) expected for nonrotating BHs, based on their innermost stable circular orbit (Rees 1984). Assuming that the seeds of high-z SMBHs have a similar radiative efficiency, and that their accretion obeys the corresponding Eddington limit, a 100 $M_\odot$ Pop III seed BH would need to accrete for $\approx$0.8 Gyr to reach 10^9 $M_\odot$. This is comparable with the age of the Universe at $z \approx 6$ and requires a duty cycle of near-Eddington accretion $f_{\rm duty} \approx 1$ over eight orders

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

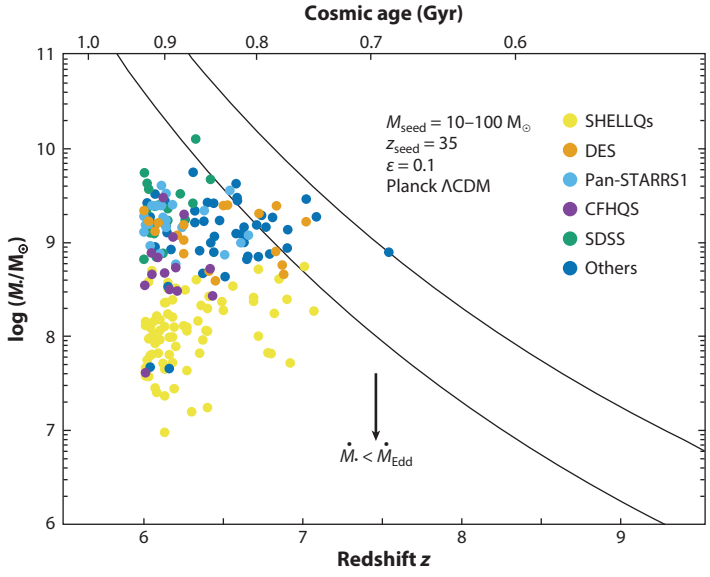

Figure 1

Masses and redshifts of quasars known at $z \geq 6$ to date. Of the 196 quasars shown, 78 sources were adopted from the compilation by Bañados et al. (2016) complete as of July 2016. Subsequently discovered quasars were added from several optical (Pan-STARRS, DES, DELS, SHELLQs) and IR (UKIDSS, VISTA) surveys, often used in combination, and often also including IR data from 2MASS and WISE. The full list is provided in the **Supplemental Table**. Masses were estimated from the rest-frame UV luminosity (M_{1450}) and assumed a constant bolometric correction and Eddington ratio ($f_{Edd} = 1$), except for the strongly lensed $z = 6.51$ quasar (Fan et al. 2019b) for which we adopted the published virial mass, including a magnification factor of 51.3. Many of the least luminous quasars, discovered predominantly in the SHELLQs survey (shown in *yellow*), have Eddington ratios below unity; the masses for these least luminous sources are underestimated by the assumption of $f_{Edd} = 1$. The pair of black curves show the mass of a BH, for reference, that grows continuously at the Eddington rate, with a radiative efficiency of $\epsilon = 0.1$, starting from a stellar-mass seed BH of $M_\bullet = 10\ M_\odot$ (*lower curve*) or $100\ M_\odot$ (*upper curve*) at $z = 35$, in a flat concordance cosmology with $\Omega_\Lambda = 0.69$, $\Omega_m = 0.31$, and $h = 0.68$ (Planck Collab. 2018). Abbreviations: 2MASS, Two-Micron All Sky Survey; BH, black hole; CFHQ, Canada-France High-redshift Quasar Survey; DELS, DESI (Dark Energy Spectroscopic Instrument) Legacy Imaging Surveys; DES, Dark Energy Survey; Pan-STARRS, Panoramic Survey Telescope & Rapid Response System; SDSS, Sloan Digital Sky Survey; SHELLQs, Subaru High-z Exploration of Low-Luminosity Quasars; UKIDSS, UKIRT Infrared Deep Sky Survey; UKIRT, United Kingdom Infrared Telescope; VISTA, Visible and Infrared Survey Telescope for Astronomy; WISE, *Wide-field Infrared Survey Explorer*.

Supplemental Material >

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

of magnitude growth in mass. Several effects make such a high duty cycle for a Pop III seed, sustained over orders of magnitude growth in mass, unlikely, including feedback from accretion onto the BH itself, as well as displacement of the gas reservoir by UV radiation and supernova (SN) explosions of the Pop III stars in the shallow gravitational potential wells of minihalos (Johnson & Bromm 2007, Whalen et al. 2008, Milosavljević et al. 2009b, Alvarez et al. 2009; see Section 3.2).

Several different scenarios have been put forward to ease these timescale constraints and help explain the existence of $M_\bullet \sim 10^9\ M_\odot$ at $z = 6$–7. Generally, the two options are to increase either the seed BH mass or the growth rate. Before enumerating these, it is worth making a few points. First, even in models with massive BH seeds, a high duty cycle is required if accretion is Eddington limited (Tanaka & Haiman 2009). Although feedback effects in minihalos make this unlikely, such efficient accretion may be easier to maintain for larger seeds residing inside more massive halos (Di Matteo et al. 2008; see Section 5). Second, it is worth emphasizing that even

for the most massive and highest-redshift SMBHs, the time-averaged accretion rate needs to be only modestly (about two times) above the Eddington-limited rate for most BH seeding models. Moderately supercritical rates, at a few times the Eddington-limited value, could be maintained with duty cycles of ∼20–30% in some accretion disk models (e.g., Sądowski 2009, Madau et al. 2014). Finally, we emphasize that only a tiny minority of early BHs, born in highly biased regions of the Universe, grow to $\sim 10^9\ M_\odot$ by $z \gtrsim 6$. The vast majority of massive BHs, born in more typical regions, remain far below this mass by this redshift.

1.3. Accelerated Growth by "Mergers and Acquisitions"

We briefly enumerate several pathways to accelerate the assembly of massive BHs in high-z galaxies. We discuss each pathway in detail in Sections 3 and 5 below.

One possibility is that a small fraction of lucky early Pop III seeds is able to sustain Eddington accretion over most of the history of the Universe. Although this is unlikely for stellar-mass seeds formed in the shallow potential wells of minihalos owing to negative feedback processes, it may be possible in rare massive halos with $\gtrsim 10^8\ M_\odot$ that form at redshifts as high as $z \gtrsim 30$ (Tanaka 2014). Alternatively, BH assembly could be accelerated if a significant fraction of the growth was due to mergers of compact objects (mostly other BHs). However, a possible issue with this scenario is that BH mergers can lead to gravitational wave (GW)-induced kicks that remove BHs from their reservoirs of dense gas (Haiman 2004).

Another, popular class of scenarios relies on the formation of $\approx 10^5\ M_\odot$ "massive seed" BHs, giving them a head start toward the SMBH regime. Though there are a number of variations, these models generally invoke the rapid collapse of chemically pristine primordial gas in so-called atomic-cooling halos (ACHs) with virial temperature $T_{\rm vir} \sim 10^4$ K. This is thought to form a $10^{5-6}\ M_\odot$ supermassive star (SMS), which then promptly collapses to a BH with a similar mass.[1] The key ingredient is the large accretion rate of the protostar, which requires the collapsing gas to remain warm ($\gtrsim$5,000 K). This in turn requires avoiding efficient metal or H_2 cooling and fragmentation. Several mechanisms/environments have been proposed to lead to this thermodynamical state, including exposing the ACH to intense Lyman–Werner (LW) radiation that photodissociates H_2 and suppresses its cooling (Omukai 2001, Oh & Haiman 2002, Bromm & Loeb 2003, Fernandez et al. 2014) and/or heating the gas in ACHs with a violent, rapid merger history (Yoshida et al. 2003, Wise et al. 2019); delaying H_2 cooling due to unusually high baryon–dark matter streaming velocities (Tanaka & Li 2014, Hirano et al. 2017); high-velocity collisions of two halos near the atomic-cooling threshold (Inayoshi et al. 2015); or some combination of these effects.

An idea related to the above scenario is that some Pop III remnant BHs find themselves at the center of an ACH without prior star formation, because of the above peculiar mechanisms/environments. Some of these BHs would then accrete at super/hyper-Eddington accretion rates and grow to $\sim 10^{5-6}\ M_\odot$ (Pacucci et al. 2015b, Inayoshi et al. 2016, Ryu et al. 2016).

A different possibility is that a $M_\bullet \approx 10^{3-4}\ M_\odot$ intermediate-mass BH (IMBH) forms promptly through stellar mergers in the core of an ultradense stellar cluster with stellar mass of $\sim 10^5\ M_\odot$ in a metal-poor protogalaxy (Omukai et al. 2008, Devecchi & Volonteri 2009). Direct collisions can occur on a timescale shorter than the lifetime of massive stars (Katz et al. 2015, Yajima & Khochfar 2016, Sakurai et al. 2017), especially if the cluster is still embedded in dense gas, where some of the protostars are accreting at high rates and have bloated envelopes, significantly increasing their geometric cross-section (Reinoso et al. 2018, Tagawa et al. 2019).

Atomic-cooling halos (ACHs): a dark matter halo with virial temperature high enough to enable radiative cooling due to atomic hydrogen transitions ($T_{\rm vir} \approx 10^4$ K)

Supermassive star (SMS): a very massive, $\sim 10^{5-6}\ M_\odot$, star predicted to form through rapid collapse and accretion of primordial gas

Lyman–Werner (LW) radiation: soft-UV photons that can dissociate H_2 via resonant absorption in electronic Lyman and Werner lines followed by a radiative cascade

Baryon–dark matter streaming velocity: relative velocity between gas and dark matter due to interactions with the radiation field before recombination

Super/hyper-Eddington accretion: accretion at a rate higher than the Eddington limit (more than an order of magnitude higher in the hyper-Eddington case)

Intermediate-mass black hole (IMBH): black hole with mass in the range of $M_\bullet = 10^{2-5}\ M_\odot$

[1] These are often referred to as direct collapse black hole (DCBH) models, although arguably this is a misnomer given the inevitable intermediate stage of an SMS. In this review, we therefore do not employ the otherwise very popular DCBH terminology.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Finally, in the absence of a definitive conclusion for early massive BH formation via the astrophysical scenarios above, it is worth keeping in mind other, more exotic possibilities. These include primordial BHs (PBHs) formed soon after the Big Bang, SMSs sustained by DM annihilation, efficient energy dissipation of magnetic fields in ACHs, and BHs fueled by self-interacting DM (SIDM; see Section 6).

1.4. Below the Tip of the Iceberg

The current surveys of distant quasars can only detect unusually bright and massive BHs that accrete near the Eddington limit. This makes the $\sim 10^9\ M_\odot$ SMBHs ultrarare objects, unrepresentative of the underlying massive BH population. Their hosts are also very massive, highly evolved galaxies (see Section 2 below), which themselves must have formed in highly biased regions of the Universe. Of course, explaining the existence of such extreme objects is crucial to improve our understanding of BH and galaxy formation in the early Universe. However, it is arguably even more important to understand the much larger population of massive BHs, currently hidden from our view because of their lower masses and/or lower accretion rates. Many early BHs can also remain undiscovered because obscuration by large amounts of gas and dust makes them too dim. This obscuration may be caused by material in the nuclear region, very close to the black hole, and/or more distant material in the host galaxy (e.g., Buchner & Bauer 2017). As discussed in Section 7, the nature of this population of massive BHs needs to be investigated to better constrain theoretical assembly models with ongoing and future observational programs. In this review, we focus on a theoretical framework of massive BH formation and growth processes for a wide range of initial BH masses, $10 \lesssim M_\bullet/M_\odot \lesssim 10^6$, addressing the formation of the relatively typical BH population as well as that of the extreme BHs.

This review is organized as follows. In Section 2, we summarize high-z quasar observations, including their current status and the most recent discoveries. In Section 3, we discuss the timescales for BH growth, taking into account the physics of accretion flows over a wide range of spatial scales and summarizing the results of radiation hydrodynamic (RHD) simulations conducted in the past decade. We then specialize to applications to the high-z Universe. In Section 4, we discuss the physics of transferring angular momentum, which is one of the biggest obstacles to maintaining rapid inflows from galactic scales down to the nuclear BH. We then again specialize to the high-z Universe. In Section 5, we summarize possible formation pathways of massive seed BHs in high-z protogalaxies, and discuss their subsequent growth, as well as the evolution of the overall population of massive BHs in the early Universe. In Section 6, we briefly mention several more exotic ways of producing massive BHs in the early Universe. Finally, we summarize future observational diagnostics of the formation and growth processes of the early massive BH population in Section 7 and offer our main conclusions in Section 8.

Many previous reviews have addressed various aspects of the above topics, including Haiman & Quataert (2004), Volonteri (2010, 2012), Volonteri & Bellovary (2012), Haiman (2013), Johnson & Haardt (2016), Latif & Ferrara (2016), Gallerani et al. (2017), Valiante et al. (2017), and Woods et al. (2019), as well as the book edited by Latif & Schleicher (2018). Here, we aim to provide a comprehensive but concise up-to-date review, focusing especially on the physics of BH formation and growth processes.

2. OBSERVATIONS

The first quasars were identified as quasi-stellar radio sources in radio surveys in the 1950s. Based on its optical spectrum, the radio source 3C 273 was interpreted as the bright nucleus of a galaxy at redshift $z = 0.158$ (Schmidt 1963). The large energy output and short time-scale variability

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

soon led to the consensus that quasars are powered by massive BHs via an accretion disk (see Rees 1984 for an early review).

Over several decades, increasingly large surveys (mainly in the optical, but also in X-ray and radio bands) mapped out the LF of quasars. These have revealed a clear evolution over cosmic time, with quasar activity rising from early times, peaking around $z \approx 2$, and falling again toward $z = 0$.

This behavior is broadly consistent with a cosmological picture in which massive BH seeds grow primarily during brief episodes of accretion. These episodes are expected to be often triggered by major mergers of their parent halos (e.g., Kauffmann & Haehnelt 2000, Volonteri et al. 2003, Hopkins et al. 2008, Somerville et al. 2008), for which there is some observational support (see Goulding et al. 2018 and references therein). Furthermore, assuming a radiative efficiency of ϵ, i.e., that the accretion rate $\dot{M}$ produces a luminosity $L = \epsilon \dot{M} c^2$ and a growth of the BH's rest mass at a rate of $(1 - \epsilon)\dot{M}$, the above picture directly links quasar activity to the local population of remnant BHs (Lynden-Bell 1969, Soltan 1982). The total quasar light output, measured by integrating the quasar LF over luminosity and redshift, is consistent with the local nuclear BH mass density of $\approx 4 \times 10^5$ $M_\odot$ Mpc^{-3}, measured using correlations between BH mass and global galaxy properties (Kormendy & Ho 2013) and an average radiative efficiency of $\epsilon \approx 10\%$ (with the latter depending on luminosity; Yu & Tremaine 2002, Shankar et al. 2004). What this broad picture is missing is where the BHs with masses of $\approx 10^{5-6}$ $M_\odot$, corresponding to the low-mass end of the SMBH mass function, come from. Over the past two decades, beginning with discoveries of distant quasars at $z \gtrsim 6$ in the SDSS, it has become clear that such seeds must have appeared very early on.

2.1. High-Redshift ($z \gtrsim 6$) Quasar Surveys

Searching for high-z quasars presents some technical challenges. Because bright quasars detectable at large redshifts are rare, large fractions of the sky need to be surveyed. The primary means of identifying quasars is based on their multicolor broad-band photometry, which allows efficient separation from the stellar locus in color space, particularly via the prominent Lyman-α break (e.g., Warren et al. 1987). At high redshifts, this requires photometry at the reddest optical bands. For example, the Lyman break falls at the center of the common optical u, g, r, i, and z band filters at $z = 1.9$, 2.9, 4.1, 5.3, and 6.5, respectively. Finally, the large amount of data requires efficient automated data processing.

These criteria were first met by the SDSS, resulting in the first handful of quasars at $z \gtrsim 6$, beginning with Fan et al. (2000). Large optical and IR surveys have continued to dominate high-z quasar searches in the past two decades (see **Table 1**). The SDSS (Jiang et al. 2016) and the CFHQS (Willott et al. 2007, 2010b) have together found several dozen quasars out to $z \lesssim 6.5$, limited by their reddest bands. The addition of a y filter extends this redshift range to $z \sim 7.2$ and has resulted in discoveries of many of the highest-z quasars by the Pan-STARRS1 survey (Bañados et al. 2016, Chambers et al. 2016, Koptelova et al. 2017, Mazzucchelli et al. 2017, Tang et al. 2017), the SHELLQs survey (Matsuoka et al. 2016, 2018a,b, 2019b), and the Dark Energy Survey (DES) (Reed et al. 2015, 2017, 2019; Yang et al. 2019), and by Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys (DELS; Wang et al. 2018, 2019; Yang et al. 2019), combined with near- and mid-IR data from several other surveys. In this combination, the quasar is a drop-out in the optical bands, but detected in the IR. Many of the highest-z sources have indeed been recently discovered by such combinations from multiple surveys, which included the UKIDSS (Lawrence et al. 2007, Mortlock et al. 2011), the UKIRT Hemisphere Survey (UHS; Wang et al. 2018), the VISTA (Visible and Infrared Survey Telescope for Astronomy) Kilo-degree Infrared Galaxy Survey (VIKING; Edge et al. 2013, Venemans et al. 2013), VLT (Very Large Telescope) Survey Telescope ATLAS (VST ATLAS; Carnall et al. 2015, Chehade et al. 2018), and the

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Table 1 List of surveys utilized in the discoveries of high-z quasars at redshift $z \geq 6$[a]

Name	Bands	Area (deg^2)	Number of quasi-stellar objects	References
Subaru (including SHELLQs + Subaru SC)	Optical g, r, i, z, y	1,400	78	SHELLQS: Matsuoka et al. 2016, 2018a,b, 2019a
	Optical z_B, z_R	7	2	Subaru SC: Kashikawa et al. 2015
Pan-STARRS1	Optical g, r, i, z, y	31,000	44	Chambers et al. 2016
DELS (including DECaLS, BASS, MzLS)	Optical g, r, z	14,000	27	Dey et al. 2019
DES (including DES SV, Yr1, and DR1)	Optical g, r, i, z, Y	5,000	18	DES Collab. et al. 2005
SDSS	Optical u, g, r, i, z	15,000	26	York et al. 2000
CFHQS (including other CFHTLS)	Optical g, r, i, z	500	15	Willott et al. 2007, 2010b
UKIDSS (including ULAS, UKIDSS-DXS, and UHS)	IR z, Y, J, H, K	7,000[b]	64	Lawrence et al. 2007
VISTA (including VHS and VIKING)	IR J, Ks	20,000	62	VHS: McMahon et al. 2013
	IR z, Y, J, H, K	1,500	31	VIKING: Edge et al. 2013; Venemans et al. 2019
VST ATLAS	Optical u, g, r, i, z + IR	4,700	4	Shanks et al. 2015
FIRST + NDWFS + FLAMEX	21 cm + optical + IR	4	1	McGreer et al. 2006
WISE (including unWISE + AllWISE)	mid-IR	All sky	71	Wright et al. 2010
2MASS	IR J, H, Ks	All sky	26	Skrutskie et al. 2006

Abbreviations: 2MASS, Two-Micron All Sky Survey; AllWISE, program combining data from the cryogenic and postcryogenic survey phases of WISE; BASS, Beijing-Arizona Sky Survey; CFHQS, Canada-France High-redshift Quasar Survey; CFHTLS, Canada-France-Hawaii Telescope Legacy Survey; DECaLS, Dark Energy Camera Legacy Survey; DELS, DESI (Dark Energy Spectroscopic Instrument) Legacy imaging Surveys; DES, Dark Energy Survey; DR1, data release 1; DXS, Deep Extragalactic Survey; FIRST, Faint Images of the Radio Sky at Twenty centimeters; FLAMEX, FLAMINGOS (Florida Multi-object Imaging Near-IR grism Observational Spectrometer) Extragalactic Survey; MzLS, Mayall z-band Legacy Survey; NDWFS, NOAO Deep Wide-Field Survey; Pan-STARRS1, Panoramic Survey Telescope & Rapid Response System 1; SDSS, Sloan Digital Sky Survey; SHELLQs, Subaru High-z Exploration of Low-Luminosity Quasars; SV, science verification; UHS, UKIRT Hemisphere Survey; UKIDSS, UKIRT Infrared Deep Sky Survey; UKIRT, United Kingdom Infrared Telescope; ULAS, UKIRT Large Area Survey; unWISE, catalog containing positions and fluxes of ~2 billion objects observed by WISE; VHS, VISTA Hemisphere Survey; VIKING, VISTA Kilo-degree Infrared Galaxy survey; VISTA, Visible and Infrared Survey Telescope for Astronomy; VST ATLAS, VLT (Very Large Telescope) Survey Telescope ATLAS survey; WISE, *Wide-field Infrared Survey Explorer*; Yr1, year 1.

[a]Some quasars were discovered independently in more than one survey, and most quasar discoveries have made use of more than one survey in their selection procedure. The fourth column includes the total number of discoveries in which data from each survey was utilized, but we note that IR surveys have typically been used to supplement those in the optical. See also the **Supplemental Table** for more details on each of the 196 quasars.

[b]The UHS extends UKIDSS to 17,900 deg^2 in the J band (Dye et al. 2018).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Supplemental Material >

VISTA Hemisphere Survey (VHS; McMahon et al. 2013, Venemans et al. 2015, Pons et al. 2019). Many discovery papers have also made use of the Two-Micron All Sky Survey (2MASS; Skrutskie et al. 2006) and the *Wide-field Infrared Survey Explorer* (WISE; Wright et al. 2010) mid-IR survey. The only $z \gtrsim 6$ quasar whose discovery involved other wavelengths is a radio-loud quasar at $z = 6.1$, found by matching optical with radio data in the FIRST survey [McGreer et al. 2006;

Table 2 List of $z \geq 7$ quasars

Name	Surveys	Redshift	$M_\bullet/M_\odot$[a]	$f_{\rm Edd}$	Reference
ULAS J1342+0928	WISE/DELS/ UKIDSS	7.541 [CII]	$7.8^{+3.3}_{-1.9} \times 10^8$	$1.5^{+0.5}_{-0.4}$	Bañados et al. 2018
HSC J1243+0100	SHELLQs	7.07 MgII	$3.3^{+2.0}_{-2.0} \times 10^8$	$0.34^{+0.2}_{-0.2}$	Matsuoka et al. 2019b
ULAS J1120+0641	UKIDSS SDSS	7.085 SiIII/CIII]/MgII	$2.0^{+1.5}_{-0.7} \times 10^9$	$1.2^{+0.6}_{-0.5}$	Mortlock et al. 2011
DELS J0038-1527	DELS/WISE/ Pan-STARRS1	7.021 MgII/OIII	$1.33^{+0.25}_{-0.25} \times 10^9$	$1.25^{+0.19}_{-0.19}$	Wang et al. 2018
DES J0252-0503	DES/VHS/ULAS/ WISE/VIKING	7.021 Lyα/NV	$\sim 1.6 \times 10^9$	Unknown	Yang et al. 2019
HSC J2356+0017	SHELLQs	7.01 Lyα	$\sim 5.5 \times 10^8$	Unknown	Matsuoka et al. 2019a

[a]All masses are published estimates based on the MgII line width and a virial mass estimator (Vestergaard & Osmer 2009), except for DES J0252-0503 and HSC J2356+0017, for which we use the proxies from the rest-frame UV luminosity (M_{1450}), assuming a constant bolometric correction and Eddington ratio. Redshifts are based on the metal and/or Lyα lines listed in the third column.

although this quasar, too, was independently discovered in optical-only (Stern et al. 2007)]. **Table 1** summarizes results from these surveys, and **Figure 1** shows the redshifts and inferred masses[2] of the 196 currently known quasars at $z \geq 6$. The full list is provided in the **Supplemental Table**.

Supplemental Material >

At the time of this writing, only six quasars are known at $z > 7$ (listed in **Table 2**), which will surely soon change with large forthcoming IR surveys, such as *Euclid* and *Wide Field Infrared Survey Telescope* (WFIRST).

Among these discoveries, the SHELLQs survey stands out as being deeper than other large–solid angle optical surveys and, therefore, able to find less-luminous quasars. This has allowed a determination of the $z \sim 6$ LF over an unprecedentedly broad range of luminosities (Matsuoka et al. 2018c) and has led to the important finding that these somewhat fainter objects have Eddington ratios that are typically lower than unity. Onoue et al. (2019) measured virial masses from the MgII line in deep optical spectra for six of the least luminous SHELLQs quasars at $6.1 \lesssim z \lesssim 6.9$. They found $f_{\rm Edd} \approx 1.1$ for one source, but $0.16 \leq f_{\rm Edd} \leq 0.43$ for the other five. This appears significantly lower than the Eddington ratios $f_{\rm Edd} \sim 1$ typically found in the past for more luminous quasars at these redshifts, and also somewhat lower than measured previously for 10 faint CFHQS quasars (Willott et al. 2010a). Shen et al. (2019) recently presented virial masses for a sample of 50 $z > 5.7$ quasars with a range of luminosities and found a median value of $f_{\rm Edd} \sim 0.3$ and Mazzucchelli et al. (2017) found an average $f_{\rm Edd} \sim 0.4$ at $z > 6.5$. Trakhtenbrot et al. (2017) also estimated the radiative efficiency of the nuclear accretion disk for the highest-redshift quasars to be $\epsilon \sim 0.03$–0.3. These recent results together suggest that the global Eddington ratio distribution at $z \sim 6$ is broader than previously measured, similar to that of low-z quasars, and with many of the most massive $\sim 10^9$ $M_\odot$ SMBHs at $z > 6$ either accreting or shining at sub-Eddington rates.

The most striking feature of the high-z quasars is their large BH masses. The masses are estimated based on virial relations (Vestergaard & Osmer 2009), typically using the width Δv of the CIV or MgII line. Inferring the BH mass from a virial relation of the form $\Delta v^2 \sim GM/r$ requires

[2]Virial or other mass estimates have only been published for a fraction of the quasars, but absolute rest-frame 1450 Å magnitudes (M_{1450}) are available for all quasars. We therefore obtained masses by assuming a constant (product of the) bolometric correction and the Eddington ratio: $\log_{10}(M_\bullet) = (-M_{1450} - 3.46)/2.5$, which yields, on average, the published virial mass estimates.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

knowledge of the size r of the broad line region (BLR), which, in turn, is calibrated on low-z quasars. Such calibrations are performed with quasars less luminous than their high-z counterparts (e.g., Peterson 2006). Thus, the high-z mass estimates rely on extrapolating these relations in both redshift and luminosity. This is somewhat justified by the fact that apart from much stronger absorption from the intergalactic medium (IGM), the spectra of the $z \sim 6$ quasars are indistinguishable from low-z quasars (Fan 2006; but see details below). This similarity, particularly in the line-to-continuum ratios, also makes it implausible that high-z quasars are preferentially beamed toward us (resulting in overestimates of their luminosities). In fact, the lack of obvious spectral differences, which include both the shape of the continuum and the strength of metal lines, more generally implies that the birth environments of quasars are established very early on (Shen et al. 2019).

One obvious question is whether luminosities (and thus masses) may have been overestimated significantly due to gravitational lensing. Strong lensing along a random line of sight to $z \sim 6$ is a priori very unlikely (probability of $\sim 10^{-3}$), but if the intrinsic (unlensed) $z \sim 6$ quasar LF is steep and/or extends to faint magnitudes, then magnification bias can boost the probability of strong lensing to even order unity (Comerford et al. 2002, Wyithe & Loeb 2002). In most cases, strong lensing would produce two detectable and resolvable images at the sensitivity and spatial resolution of the *Hubble Space Telescope* (HST; Keeton et al. 2005), yet a search among ∼200 quasars has not revealed any multiple images (Richards et al. 2006, McGreer et al. 2014). This would naively rule out the possibility that most $z \sim 6$ quasars are strongly lensed. However, a strongly lensed quasar has recently been discovered at $z = 6.51$, with three images and an inferred total magnification of a factor of ≈50 (Fan et al. 2019b). This source was lensed by an unusually faint foreground galaxy, whose starlight did not significantly contaminate the quasar's spectrum. Intriguingly, this suggests that for a typical, brighter lens, the background lensed quasar would not be identified as a quasar by traditional color-selection criteria. A significant population of strongly lensed, high-redshift quasars could therefore still be missing from the existing surveys (Fan et al. 2019b, Pacucci & Loeb 2019).

2.2. Properties of the $z \sim 6$ Quasar Population

LFs have been measured in several surveys (Willott et al. 2010b, Jiang et al. 2016), with the most complete determination extending to the lowest luminosities by the SHELLQs project on the Subaru telescope (Matsuoka et al. 2018c). As shown in **Figure 2**, the LF at $z = 6$ is roughly consistent with an extrapolation from lower redshifts, with its shape remaining self-similar (well-fit by a broken power law), but the normalization dropping steeply with redshift, with quasars at $z \approx 6$ about 100 times less abundant than they are at $z \approx 4$. One important finding from the SHELLQs project is that the quasar LF flattens significantly toward lower luminosities so that the total (faint+bright) quasar population could not provide enough photons to keep the IGM ionized (see also Jiang et al. 2008), even assuming a clumping factor of ≈1 (Madau et al. 1999). This indicates that quasars are not a major contributor to cosmic reionization.

However, it is also important to consider how complete the current quasar surveys may be, as we could be missing many lensed quasars (see above) as well as a population of heavily obscured quasars. The latter can be an especially large effect, because the optical selection of the current $z > 6$ quasars is highly biased against obscured quasars. The first and only known highly obscured quasar candidate at this redshift, detected in X-rays (Vito et al. 2019), is optically classified as a Type 1 AGN. A comparison of the extrapolation of the $z = 3$–6 X-ray AGN LF (which does not select against obscured quasars) with the optically selected $z > 6$ LF (Matsuoka et al. 2018c) suggests that as much as 80–90% of all $z > 6$ quasars may be obscured and missed by optical/IR surveys (Vito et al. 2018).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

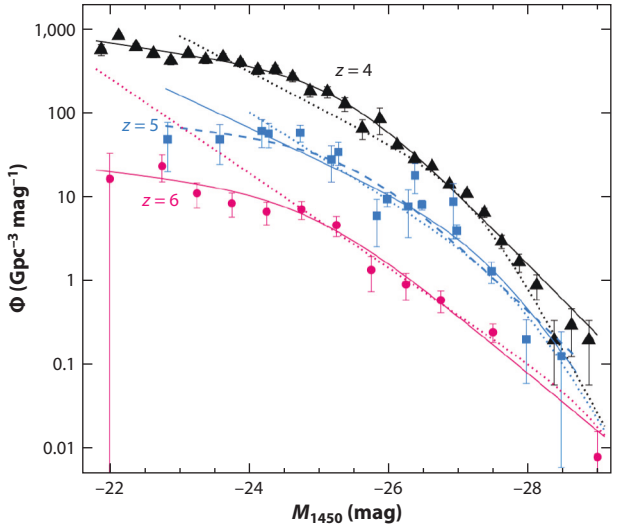

Figure 2

High-z quasar luminosity functions from SHELLQs ($z = 6$, *red circles*; Matsuoka et al. 2018c), CFHTLS ($z = 5$, *blue squares*; McGreer et al. 2018), and a combination of data from SDSS and Subaru's Strategic Program Wide survey ($z = 4$, *black triangles*; Akiyama et al. 2018). The dotted lines show broken power-law fitting functions. Abbreviations: CFHTLS, Canada-France-Hawaii Telescope Legacy Survey; SDSS, Sloan Digital Sky Survey; SHELLQs, Subaru High-z Exploration of Low-Luminosity Quasars. Figure adapted with permission from Matsuoka et al. (2018c).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Overall, the high-redshift quasars individually look very similar to their low-z counterparts. In particular, the hosts of the $z \gtrsim 6$ quasars contain significant amounts of metals and dust, including the host of the most-distant quasar at $z = 7.54$ (Venemans et al. 2017b, Novak et al. 2019). Copious amounts of metals are revealed by observations of molecular lines (e.g., CO, [CII]), in the ISM of the hosts on kiloparsec scales (Bertoldi et al. 2003; Walter et al. 2003, 2009; Wang et al. 2013; Willott et al. 2015; Venemans et al. 2017a, 2019). The amount of cool molecular gas is $\sim 10^{9–10}$ $M_\odot$ (Carilli & Walter 2013). The far-IR continuum in these observations likewise reveals a great amount ($\sim 10^{7–8}$ $M_\odot$) of warm, thermally emitting dust.

The highest-angular resolution observations by instruments such as Atacama Large Millimeter Array (ALMA) and Institut de Radio Astronomie Millimétrique (IRAM) have spatially resolved the hosts of many luminous high-redshift quasars and found a diverse range, which include compact dispersion-dominated systems and rotationally supported galaxies, as well as isolated galaxies, major mergers, and close companions in some cases (Decarli et al. 2017, Neeleman et al. 2019).

On larger scales, one would naively expect that the luminous quasars at any redshift should reside in the most massive halos, which are in the most overdense environments (although cosmological simulations suggest that this is not strictly true; Fanidakis et al. 2013). Several surveys have looked for a corresponding excess overdensity of galaxies around high-z quasars on megaparsec scales. However, the evidence is inconclusive: The environments of some of the quasars show galaxy overdensities, and some do not (Kim et al. 2009, Utsumi et al. 2010, McGreer et al. 2014, Balmaverde et al. 2017, Mazzucchelli et al. 2017, Ota et al. 2018, Habouzit et al. 2019).

The overall strength and kinematics of the molecular lines and of the continuum dust emission are together consistent with these hosts being analogs of low-redshift starburst galaxies, with the

dust being heated by star formation on kiloparsec scales, at star-formation rates of up to a few 1,000 $M_\odot$ year^{-1}. Likewise, the nuclei of these hosts appear highly enriched on 0.1–1 pc scales, as evidenced by broad metal emission lines, such as CIV in their rest-frame UV spectra, which are similar to those of their low-z counterparts (De Rosa et al. 2014, Mazzucchelli et al. 2017, Reed et al. 2019, Shen et al. 2019). The FeII/MgII line ratio, a proxy for the chemical abundance of the BLR gas in bright quasar hosts, also shows no redshift evolution (De Rosa et al. 2011). The CIV lines do show unusually large blueshifts, indicating that winds driven out of the nuclear disks may be especially strong in the luminous, high-z quasars. In addition to the systematically larger blueshifts of broad emission lines at higher redshift (Meyer et al. 2019), there have been tentative claims for a systematically larger fraction of weak-line quasars at $z > 6$ (Bañados et al. 2016, Meyer et al. 2019, Shen et al. 2019). In total, the evidence above points to an early, rapid assembly of the massive host galaxies of the highest-redshift quasar BHs.

An interesting tentative difference in the hosts of the highest-z quasars is the ratio of their BH and galaxy masses. The resolved kinematics of the ISM, especially from the strongest [CII] emission line, tend to yield dynamical masses for the fainter, lower-luminosity quasars that obey local scaling relations, albeit with a larger scatter (Willott et al. 2017, Izumi et al. 2018). However, the hosts of more luminous high-z quasars appear to have dynamical masses an order of magnitude below the corresponding low-z relations (Wang et al. 2016b, Decarli et al. 2018, Shimasaku & Izumi 2019). If confirmed, this suggests that the most massive SMBHs at $z \gtrsim 6$ got a head start over the growth of their host galaxies, which is perhaps in slight tension with the high metal and dust enrichment of these hosts. The clear caveats are that gas tracers can underestimate dynamical masses and that the brightest quasi-stellar objects (QSOs) can suffer from a selection bias that picks out preferentially massive BHs (see Volonteri & Stark 2011, Lupi et al. 2019 and references therein for discussions of such biases). We also note that the SMBHs at the high end of the locally measured $M_\bullet$–$\sigma_\star$ relation also tend to have higher masses, but there is a similar tentative upturn in the Faber–Jackson relation between host luminosity $L_{\rm gal}$ and $\sigma_\star$, so that the most massive BHs are not outliers in the M–$L_{\rm gal}$ relation (Lauer et al. 2007). It is possible that high-z quasars fit the same trend. However, we emphasize that the local relations relate BH masses to properties of the bulge component, about which we have no information at high redshift. More generally, a key missing piece of evidence is the direct observation of starlight from the high-z quasar host galaxies at rest-frame UV to near-IR wavelengths (Fan et al. 2019a). The one exception is the UV starlight detected in a $z = 6.2$ system (Decarli et al. 2019), interpreted to be a merger.

Finally, an interesting constraint can be placed on the growth of high-redshift quasars from the size of their cosmological ionized regions. At high redshift, when the IGM within a few (physical) megaparsecs of the quasar is substantially neutral, the apparent size of the ionized region can be used to probe the neutral fraction of the local IGM, as well as the quasar's luminosity-weighted age (Bolton & Haehnelt 2007, Cen & Haiman 2000, Mesinger & Haiman 2004, Wyithe & Loeb 2004). Some of the $z \gtrsim 6$ quasars appear to have very compact cosmological ionized regions, as inferred from their Lyman-α absorption spectra (Eilers et al. 2017, 2018; Davies et al. 2019), implying that they could not have grown a significant fraction of their mass at their observed luminosity. These results suggest that most of the BH mass of these quasars was assembled via either radiatively inefficient or highly obscured phase.

3. ACCRETION AND RADIATIVE FEEDBACK

In this section, we review the theoretical framework of BH accretion and discuss rates at which preexisting BHs can grow by accretion, particularly in the face of radiative feedback. This is

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

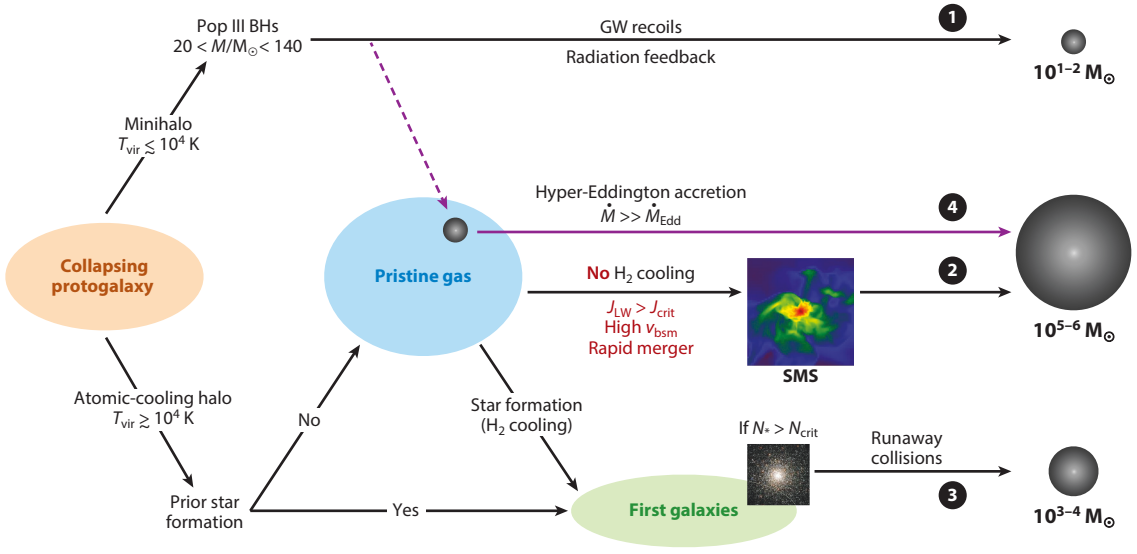

Figure 3

Formation pathways of seed BHs in early protogalaxies: ❶ Pop III remnant BHs with a mass of $M_\bullet \approx 10^{1-2}$ $M_\odot$, ❷ massive seed BHs with $M_\bullet \approx 10^{5-6}$ $M_\odot$ in ACHs under peculiar conditions such as strong LW radiation ($J_{LW} > J_{crit}$), high baryon-DM streaming velocity, and rapid mergers of DM halos, and ❸ relatively massive seeds with $M_\bullet \approx 10^{3-4}$ $M_\odot$ via runaway collisions in ultradense stellar clusters. ❹ Hyper-Eddington accretion onto stellar-mass BHs ($\dot{M}_\bullet \gg \dot{M}_{Edd}$) would effectively result in a massive seed at the center of a dense pristine gas cloud. Abbreviations: ACH, atomic-cooling halo; BH, black hole; DM, dark matter; GW, gravitational wave; LW, Lyman–Werner; Pop III, Population III; SMS, supermassive star.

motivated by the natural availability of stellar-mass BHs in the early Universe, left behind by the first generation of stars. As we argue, it is possible for BHs to grow at highly super-Eddington rates, which represents one of the pathways for rapid BH assembly in the early Universe. These pathways are illustrated in **Figure 3**, along with other possibilities that will be discussed in the sections below. However, we emphasize that no self-consistent calculation to date has included all the necessary multi-scale physics *and* followed the BH growth over several orders of magnitude in mass. We first focus on the basic underlying physics (Section 3.1) and then discuss applications to the high-z Universe (Section 3.2 and Section 3.3).

For convenience, **Figure 4** illustrates the structure of accretion flows onto a BH embedded in a protogalaxy. The characteristic physical scales and mechanisms relevant to the discussions below are listed in **Table 3** (with their definitions and fiducial values).

3.1. Growing Black Holes by Accretion: Is There an Eddington Limit?

Assuming that high-z SMBHs grow mostly via rapid gas accretion and radiate ∼10% of the rest mass energy of accreting matter, as low-z quasars do on average (Soltan 1982, Yu & Tremaine 2002, Ueda et al. 2003), the outward radiation pressure force on the infalling gas, through electron scattering, matches the inward gravitational force at the critical accretion rate of $\dot{M}_{Edd} \equiv 10\ L_{Edd}/c^2$, where $L_{Edd} = 4\pi cGM_\bullet/\kappa_{es}$ is the Eddington luminosity.[3] If accretion is limited to this rate, the

[3]This definition includes a fiducial factor of 10, which assumes a radiative efficiency of 10%. We employ this definition throughout this review, but we caution the reader that an equally common definition in the literature is $\dot{M}_{Edd} \equiv L_{Edd}/c^2$, i.e., excluding this factor.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

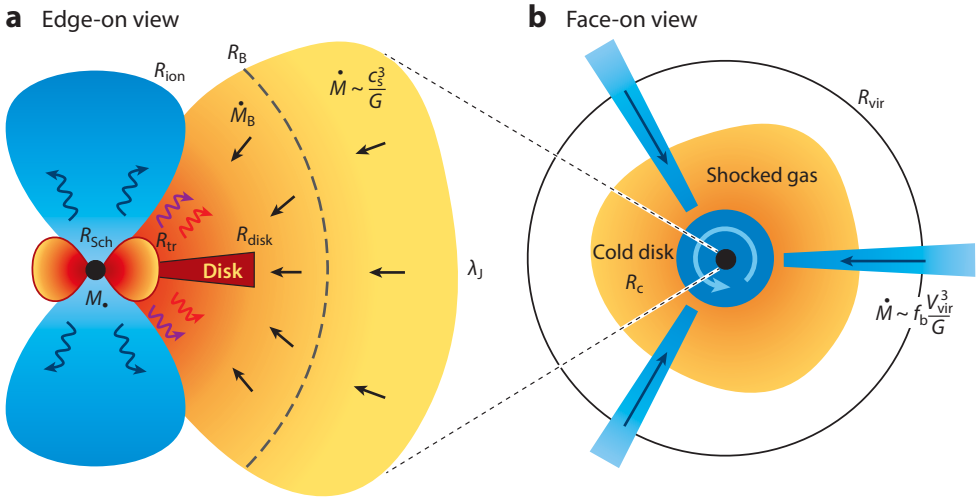

Figure 4

Schematic illustrations of the accretion flow onto a massive BH with a mass, $M_\bullet$, at a rate significantly exceeding $\dot{M}_{Edd}$ (panel *a*, edge-on view) and of the early protogalaxy that hosts the accreting BH (panel *b*, face-on view). The characteristic radii and mass accretion rates of this system, shown in the illustrations, are summarized in the accompanying **Table 3**, along with their definitions and fiducial values. Abbreviation: BH, black hole.

Table 3 List of relevant physical scales and related quantities discussed in this review

Quantity	Symbol	Approximation[a]
Jeans mass	$M_J \equiv \rho \lambda_J^3$	$2 \times 10^4 \, n_{H,4}^{-1/2} T_3^{3/2}$
Eddington accretion rate	$\dot{M}_{Edd} \equiv \frac{L_{Edd}}{0.1c^2}$	$2.3 \times 10^{-5} \, M_{\bullet,3}$
Bondi accretion rate	$\dot{M}_B \equiv \pi e^{3/2} \rho \frac{G^2 M_\bullet^2}{c_s^3}$	$4.5 \times 10^{-3} \, n_{H,4} T_3^{-3/2} M_{\bullet,3}^2$
Accretion rate in an unstable cloud	$\dot{M} \sim \frac{c_s^3}{G}$	$4 \times 10^{-3} \, T_3^{3/2}$
Mass inflow rate from galactic scales	$\dot{M} \sim f_b \frac{V_{vir}^3}{G}$	$6 \times 10^{-2} \, T_{v,4}^{3/2}$
Schwarzschild radius	$R_{Sch} \equiv \frac{2GM_\bullet}{c^2}$	$2 \times 10^{-5} \, M_{\bullet,3}$ (AU)
Photon trapping radius	$R_{tr} \equiv \frac{\kappa_{es} \dot{M}_\bullet}{4\pi c}$	$0.01 \, M_{\bullet,3} \left(\frac{\dot{m}}{100}\right)$ (AU)
Bondi radius	$R_B \equiv \frac{GM_\bullet}{c_s^2}$	$0.6 \, T_3^{-1} M_{\bullet,3}$ (pc)
Jeans length	$\lambda_J \equiv \sqrt{\frac{\pi k_B T}{G \mu m_p \rho}}$	$4 \, n_4^{-1/2} T_3^{1/2}$ (pc)
Centrifugal radius (halo scale)	$R_c \equiv \lambda R_{vir}$	$26 \, \lambda_{0.05} T_{v,4}^{1/2} \left(\frac{1+z}{16}\right)^{-3/2}$ (pc)
Halo virial radius	R_{vir}	$520 \, T_{v,4}^{1/2} \left(\frac{1+z}{16}\right)^{-3/2}$ (pc)

[a]The units for mass and accretion rate are $M_\odot$ and $M_\odot$ year^{-1}, respectively. The BH mass is $M_\bullet = 10^3 M_{\bullet,3} \, M_\odot$; gas density, $n_H = 10^4 n_{H,4}$ cm^{-3}; gas temperature, $T = 10^3 T_3$ K; DM halo virial temperature, $T_{vir} = 10^4 T_{v,4}$ K; DM halo spin parameter, $\lambda = 0.05 \lambda_{0.05}$; and $\dot{m} \equiv \dot{M}_\bullet / \dot{M}_{Edd}$ is the dimensionless BH accretion rate normalized by the Eddington rate (at 10% radiative efficiency, as defined in the second row).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

timescale for growth to the BH mass $M_\bullet$ becomes as long as

$$t_{\rm grow} \approx \frac{0.45\,\epsilon}{(1-\epsilon)f_{\rm duty}}\,\ln\left(\frac{M_\bullet}{M_{\rm seed}}\right)\ {\rm Gyr} \approx 0.81\ {\rm Gyr}, \qquad 1.$$

Photon trapping radius ($R_{\rm tr}$): the radius at which radiation is trapped in the accretion flow and advected into the black hole

where $f_{\rm duty}$ is the duty cycle of accretion, $M_{\rm seed}$ is the initial seed mass, and the last step adopts the fiducial values $\epsilon = 0.1, f_{\rm duty} = 1, M_{\rm seed} = 100\ {\rm M}_\odot$, and $M_\bullet = 10^9\ {\rm M}_\odot$. This estimate shows that the growth timescale is comparable with the age of the Universe at $z \sim 6$, even when continuous and rapid gas supply is assumed (Haiman & Loeb 2001, Madau & Rees 2001, Volonteri et al. 2003, Li et al. 2007). This, in turn, raises basic questions, such as: What is the radiative efficiency of early BHs? Can accretion occur at rates exceeding the fiducial Eddington-limited value? Can the required fuel supply be maintained over several orders of magnitude growth in mass?

3.1.1. Photon trapping on small scales near the black hole. Gas accreting onto a BH releases a great amount of energy at the vicinity of the BH event horizon, $R_{\rm Sch} \equiv 2GM_\bullet/c^2 \approx 300$ km $(M_\bullet/100\ {\rm M}_\odot)$. The intense radiation would naively limit the BH growth below the critical Eddington value. However, does the Eddington limit really matter for the accretion rate? In general, AGNs appear to obey the Eddington limit on the luminosity, based on BH masses from reverberation measurements (e.g., Peterson et al. 2004). However, some X-ray binaries, such as SS 443 in our Galaxy (e.g., Okuda 2002) and some ultraluminous X-ray sources (ULXs), which are suspected to contain stellar-mass BHs (King et al. 2001, Winter et al. 2006), are believed to accrete at super-Eddington rates (Poutanen et al. 2007, Kawashima et al. 2012). Several ULXs have been observed to pulsate on a timescale of ∼1 s, implying a stellar-mass source and, thus, favoring super-Eddington accretion (e.g., King et al. 2017, and references therein). In the SMBH regime, narrow-line Seyfert-1 galaxies are presumed to be super-Eddington accretors (Mineshige et al. 2000, Wang & Netzer 2003, Collin & Kawaguchi 2004, Du et al. 2014).

The possibility of super-Eddington accretion has been explored theoretically by many authors. A basic reason why this may be feasible goes as far back as Begelman (1979). In a spherically symmetric accretion flow at a rate of $\gg \dot{M}_{\rm Edd}$, in which the radiation pressure force is supposed to halt the inflow, the emergent radiation flux is reduced by photon trapping in the optically thick accreting matter. This trapping effect operates when the radial gas inflow speed is faster than the outward photon diffusion speed; i.e., $|v_r| > c/\tau$, where τ $(=\rho\kappa_{\rm es}r)$ is the optical depth to electron scattering. This condition is satisfied within the so-called photon trapping radius defined by

$$R_{\rm tr} \equiv \frac{\kappa_{\rm es}}{4\pi c}\dot{M}_\bullet = 5\dot{m}R_{\rm Sch}, \qquad 2.$$

where $\dot{m} \equiv \dot{M}_\bullet/\dot{M}_{\rm Edd}$. The trapping effect becomes physically relevant when this radius is outside $R_{\rm Sch}$, i.e., for $\dot{m} \gtrsim 0.2$. Because most of the radiation produced inside $R_{\rm tr}$ is advected with the flow because of electron scattering, the diffusive luminosity seen at larger radii is limited to $L \lesssim GM_\bullet\dot{M}_\bullet/R_{\rm tr} = L_{\rm Edd}$, independent of the mass inflow rate. Therefore, the BH growth rate is unlimited and can exceed the Eddington value by an arbitrary factor, as long as a correspondingly large amount of inflowing gas is maintained from larger scales down to the vicinity of the BH (see also Begelman 1978, who constructed a global spherical accretion solution for ionized gas at $\dot{M}_\bullet \gg \dot{M}_{\rm Edd}$).

The above consideration holds, however, only in spherical symmetry and ignores the question of the stability of the flow. Subsequent analytical work (e.g., Quataert & Gruzinov 2000, Blandford & Begelman 2004), as well as early multidimensional hydrodynamical simulations (e.g., Igumenshchev & Abramowicz 1999, Stone et al. 1999), including those with magnetic fields (e.g., Igumenshchev et al. 2003), suggested that when $\dot{M}_\bullet \gg \dot{M}_{\rm Edd}$, these so-called radiatively inefficient

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

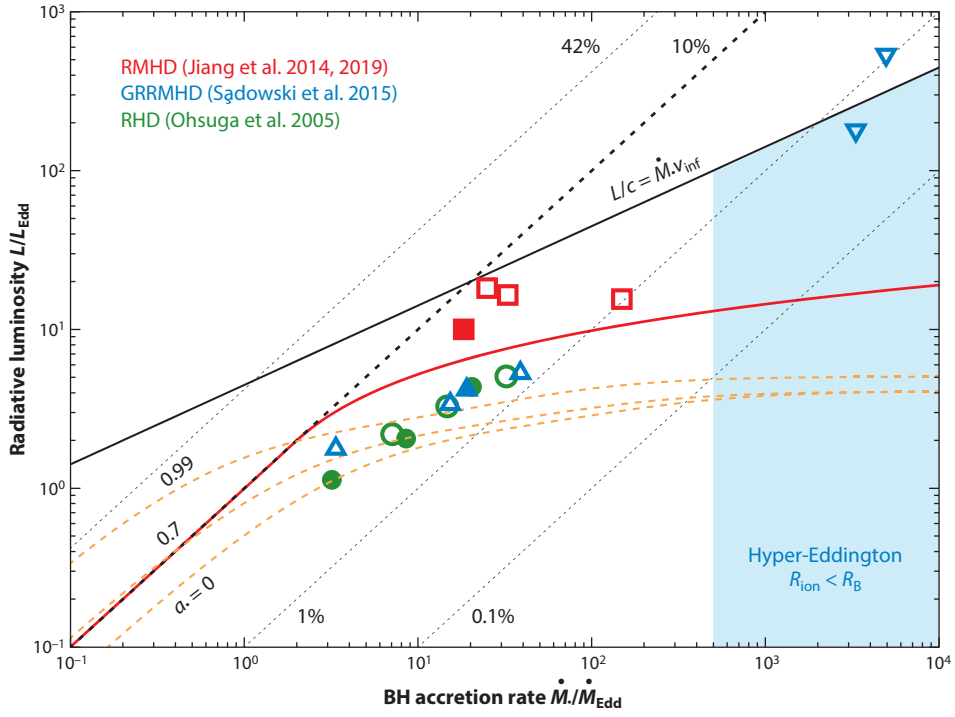

Figure 5

Summary of theoretical results for radiative luminosity versus BH accretion rate. In the analytical, slim-disk model assuming a pseudo-Newtonian potential (Watarai et al. 2000; *red line*) or taking into account the GR effect around a BH with a spin of $a_\bullet$ (Sądowski et al. 2015; *orange dashed lines*), the radiative luminosity gradually increases as $L \propto \ln(\dot{m})$ at high rates. Simulation results are shown by green circles (RHD; Ohsuga et al. 2005) for metallicities $Z = 0$ (*filled symbols*) and $Z = Z_\odot$ (*open symbols*), blue triangles (GRRMHD; Sądowski et al. 2015) for $a_\bullet = 0$ (*open*) and 0.9 (*filled*), and red squares (RMHD; Jiang et al. 2014, 2019) for stellar-mass BH (*filled*) and SMBH/AGN (*open*). For a highly magnetized accretion disk around a rapidly spinning BH $a_\bullet = 0.9$ (*inverted triangles*), the disk transits into a magnetically arrested disk state, producing higher radiative luminosity but with a lower efficiency. Metallicity and BH spin both impact the structure of the flow, via opacity and radiative efficiency near the BH, respectively. These simulations find self-consistent super-Eddington accretion on small scales with lower values of the radiative efficiency below 10% (*thick dashed black line*) but are numerically limited to model only short durations and small scales. In the shaded region, hyper-Eddington accretion from the BH sphere of influence R_B would be realized and sustained, because the ionized region is smaller than the Bondi radius ($R_{ion} \lesssim R_B$), and radiative feedback is therefore unable to suppress the inflow. The efficient growth phase can stably exist unless the outward momentum L/c dominates the inward ram pressure of the rapidly accreting gas (*solid black line*). Abbreviations: AGN, active galactic nucleus; BH, black hole; GR, general relativity; GRRMHD, general relativistic radiation magnetohydrodynamic; RHD, radiation hydrodynamic; RMHD, radiation magnetohydrodynamic; SMBH, supermassive black hole.

accretion flows (RIAFs) become unstable to outflows, and only a small fraction of the mass reaches the event horizon.

The photon-trapping effect has more recently been incorporated into accretion disk models including direct RHD simulations. In **Figure 5**, we summarize theoretical predictions of the radiative luminosity as a function of the dimensionless accretion rate captured by the BH based on both analytical work and RHD results. In the slim-disk analytical model, the radiative luminosity is proportional to $\dot{M}_\bullet$ in the sub-Eddington regime and gradually increases as $L/L_{Edd} \propto \ln(\dot{m})$ in the

super-Eddington regime (Abramowicz et al. 1988, Watarai et al. 2000). Rotating BHs produce radiation more efficiently at lower accretion rates ($\epsilon \approx 0.42$ for the dimensionless spin parameter $a_\bullet = 0.99$), but the luminosity is still saturated at $L/L_{\rm Edd} \sim 3$ at higher rates of $\dot{m} \gtrsim 10$ (Sądowski 2009; see **Figure 5** for three different spin parameters). In the past decade, RHD simulations including magnetic fields (radiation magnetohydrodynamics or RMHD) and general relativistic effects (general relativistic radiation magnetohydrodynamics or GRRMHD) have revealed the properties of rapidly accreting gas within a few 100 $R_{\rm Sch}$ of the BH (Ohsuga et al. 2005; Fragile et al. 2014; Jiang et al. 2014; McKinney et al. 2014, 2015; Sądowski et al. 2015; Takahashi & Ohsuga 2015). The radiative efficiency modestly decreases with the accretion rate down to $\epsilon \approx 1$–5% at $3 \lesssim \dot{m} \lesssim 150$.

The numerical results are overall qualitatively consistent with the analytical model but have some discrepancies. In fact, the efficiency of photon trapping is significantly reduced due to non-inflowing gas motion caused by radiation pressure and magnetic buoyancy (Jiang et al. 2014), which are not taken into account in the analytical models. Importantly, the radiative efficiency obtained in simulations with approximate numerical algorithms for radiative transfer that impose local closure relations between the radiation pressure tensor and radiation energy density, such as the flux-limited diffusion (FLD; Ohsuga et al. 2005) or the so-called M1 closure (Sądowski et al. 2015),[4] is systematically lower than those with a more accurate numerical algorithm to solve the time-dependent radiative transfer equations directly (Jiang et al. 2014, 2019). In the latter case, rapid gas accretion is still allowed in the equatorial region, but a large amount of radiation emerges with $1 \lesssim L/L_{\rm Edd} \lesssim 20$ toward the polar conical regions.

Of particular interests to us are simulations at the highest accretion rates. It would be a remarkable coincidence if the mass supply rate from large scales precisely tracked $\sim \dot{M}_{\rm Edd}$ as a BH grows by orders of magnitude in mass. More likely, the mass supply rate is initially much larger and then gradually decreases when measured in Eddington units. Jiang et al. (2019) have explored a case with $\dot{m} \approx 150$, where even the polar funnel regions become optically thick. Because the disk has inflows instead of launching strong outflows, radiation is effectively trapped and advected toward the BH, with a small fraction of the radiation able to diffuse outward, matching the expectations of the analytical models. Sądowski et al. (2015) reported a radiative efficiency as low as $\epsilon \approx 0.01$ at their very high accretion rates of $\dot{m} \approx 5 \times 10^3$. In this latter simulation, the BH was assumed to be rotating with $a_\bullet = 0.9$, and the initial poloidal magnetic field in the disk was assumed to be in a so-called magnetically arrested disk (MAD) state. In this state, the accreting gas drags the poloidal magnetic field to the center such that the accumulated, strong field disrupts the inflow structure and is likely to produce outflows and/or jets. Once the disk turns into a MAD state, the luminosity becomes as high as $\sim$100 $L_{\rm Edd}$, but the radiative efficiency is still as low as $<1\%$ even for a rapidly rotating BH (see also Narayan et al. 2003, Tchekhovskoy et al. 2011, McKinney et al. 2015).

It is worth noting that these simulations have explored the properties of accretion flows on small scales, assuming a compact torus in hydrostatic equilibrium as the initial state or adopting mass-input boundary conditions. Even though the BH feeding rate is high, as shown in **Figure 5**, a steady accretion disk forms only within $R_{\rm disk} \sim 20$–$100\ R_{\rm Sch}$, which corresponds to the (half) radius of the location of the density peak of the initial torus. Outside the steady disk, a significant fraction of the gas is ejected in a wind, and in fact the mass-loss rate dominates significantly

[4]The two approximated treatments cannot capture the angular distribution of photons near the photosphere accurately. Because the radiation flux in FLD points toward any gradient of radiation energy density, unphysical radiation flux will be produced and, thus, such a structure would likely be smeared out. In the M1 closure method, the collimation level of the radiation-driven outflow might be affected, because photons in the outflow are merged into a single beam near the photosphere.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

over the BH feeding rate. Imposing mass-input from the outer boundary, the mass inflow rate decreases toward the center owing to strong outflows ($\dot{M}_{\rm in} \propto r^s$, where $0 \lesssim s \lesssim 1$; see Blandford & Begelman 2004), and thus only a fraction $(R_{\rm disk}/R_{\rm tr})^s$ of the inflowing gas reaches a steady accretion disk and is ultimately contributing to the growth of the BH. The existence of outflows launched from a rapidly accreting BH seems ubiquitous and could potentially even reverse the inflow. Therefore, it is crucial to address how these super-Eddington accretion simulations relate to the outer boundary conditions on larger scales and to assess whether, and by how much, the radiative and mechanical outputs might suppress the gas inflow at the BH's horizon.

In summary, high accretion rates exceeding the Eddington value are possible but produce intense radiation flux toward the polar directions with $L \approx O(1\text{–}10)\ L_{\rm Edd}$. These results, however, are valid only as long as a sufficient amount of gas at rates of $\dot{M} \gg \dot{M}_{\rm Edd}$ is supplied from larger scales without being impeded by the strong radiation feedback.

3.1.2. Inflow from large scales. Gas inflows from larger scales ($r \gg R_{\rm Sch}$) can be triggered by several physical processes. First, baryons accrete into a DM halo along well-defined cold filamentary streams connected with the large-scale cosmic web. The mass accretion rate averaged over cosmological timescales is approximately given by

$$\dot{M} \approx f_{\rm b} \frac{V_{\rm circ}^3}{G} \approx 0.3\ {\rm M}_\odot\ {\rm year}^{-1} \left(\frac{V_{\rm circ}}{20\ {\rm km\ s}^{-1}} \right)^3, \qquad 3.$$

where $f_{\rm b} \approx 0.16$ is the global mean baryon fraction and $V_{\rm circ}$ is the circular velocity of the halo. Such large-scale inflows are expected to be triggered by major mergers of two galaxies (Springel et al. 2005, Hopkins & Quataert 2010, Mayer et al. 2010) or in massive DM halos in which the gas cooling timescale is significantly shorter than the dynamical timescale (Birnboim & Dekel 2003, Kereš et al. 2005, Dekel & Birnboim 2006, Di Matteo et al. 2012). Strong perturbations in both gas and stars in a merging galaxy lead to nonaxisymmetric spiral structures, which transport angular momentum and induce mass accretion down to smaller scales (see detailed discussion in Section 4 below).

Second, the rapidly accreted pristine gas settles into a compact circumnuclear disk, which becomes gravitationally unstable and thus leads to fragmentation and clump formation (Oh & Haiman 2002, Lodato & Natarajan 2006, Dekel et al. 2009). Because primordial gas is as warm as $T \sim 10^3$ K due to the absence of metal cooling (Palla et al. 1983), massive self-gravitating clumps form with a Jeans mass of $M_{\rm J} \approx 2 \times 10^4\ {\rm M}_\odot\ n_{\rm H,4}^{-1/2} T_3^{3/2}$ and collapse at rates of

$$\dot{M} \approx \frac{M_{\rm J}}{t_{\rm ff}} \approx \frac{c_{\rm s}^3}{G} \approx 4 \times 10^{-3}\ {\rm M}_\odot\ {\rm year}^{-1} T_3^{3/2}, \qquad 4.$$

where $n_{{\rm H},x} \equiv n_{\rm H}/(10^x\ {\rm cm}^{-3})$ and $T_y \equiv T/(10^y\ {\rm K})$. Note that the accretion rate depends only on the temperature and not on the density (Larson 1969, Penston 1969). The physical size of the collapsing clump is given by the Jeans length, $\lambda_{\rm J} \approx 4\ {\rm pc}\ n_{\rm H,4}^{-1/2} T_3^{1/2}$, which is substantially smaller than the halo scale but still far away from the central BH itself. If a seed BH is embedded in such an unstable cloud, the mass accretion onto the BH is much higher than the Eddington rate, namely $\dot{M}/\dot{M}_{\rm Edd} \approx 2 \times 10^3\ T_3^{3/2} (M_\bullet/100\ {\rm M}_\odot)^{-1}$.

Third, on smaller scales, the dynamics of accreting gas is finally influenced by gravity of the central BH. The characteristic scale is the so-called Bondi radius, defined by $R_{\rm B} \equiv GM_\bullet/c_{\rm s}^2 \approx 0.06\ {\rm pc}\ T_3^{-1} (M_\bullet/100\ {\rm M}_\odot)$, where $c_{\rm s}$ is the sound speed of the gas. Gas is captured by the BH and begins to accrete from the Bondi radius. If the specific angular momentum of the gas is sufficiently small, a centrifugally supported disk forms only inside the Bondi radius, and gas flows inward

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

at supersonic velocities. The characteristic accretion rate at this radius is the Bondi accretion rate,

$$\dot{M} \approx \dot{M}_{\rm B} \equiv \pi e^{3/2} \rho \frac{G^2 M_\bullet^2}{c_{\rm s}^3},$$

$$\approx 4.5 \times 10^{-3}\ {\rm M_\odot\ year^{-1}}\ n_{\rm H,6}\ T_3^{-3/2} \left(\frac{M_\bullet}{100\ {\rm M_\odot}}\right)^2, \quad 5.$$

Bondi accretion: the analytic description of spherical accretion onto a compact object traveling through a medium with uniform density and temperature

where ρ is the mass density at $r = R_{\rm B}$, the gas is assumed to be isothermal (Bondi 1952), and $\dot{M}/\dot{M}_{\rm Edd} \approx 2 \times 10^3\ n_{\rm H,6}\ T_3^{-3/2} (M_\bullet/100\ {\rm M_\odot})$. The Bondi rate should in general be considered an upper limit on the accretion rate, because it assumes free fall of gas from the Bondi radius. Negative effects associated with BH feedback, gas rotation, and MHD-winds reduce the inflow rate (Proga & Begelman 2003, Li et al. 2007, Sijacki et al. 2007, Tchekhovskoy et al. 2011, Sądowski et al. 2015), and even in the absence of these effects, when the gas is very cold, it is susceptible to gravitational perturbations that determine the inflow rate (Hopkins & Quataert 2010). Importantly, the Bondi radius is generally much larger than the trapping radius, namely $R_{\rm B}/R_{\rm tr} \approx 7 \times 10^3\ (\dot{m}/10^3)^{-1} T_3^{-1}$ (see **Figure 4**).

3.1.3. Photoionization and heating. Radiative feedback–associated BH accretion can play a crucial role on the inflow rate from the Bondi radius, where gas is only marginally bound to the BH. Even if a BH is embedded in a self-gravitating cloud, the ratio of the thermal energy to the gravitational energy is as high as $\gtrsim 5/\pi^2 \approx 0.5$ (Larson 1969, Truelove et al. 1997). Therefore, photoionization and heating by the central BH can unbind the gas and suppress gas inflow from large scales. Unfortunately, no multidimensional simulation can self-consistently resolve all the relevant scales from the event horizon to the Bondi radius. To roughly quantify this effect, let us instead approximate the size of the ionization bubble by the Strömgren radius in a uniform medium,

$$R_{\rm ion} \approx \left(\frac{3 Q_{\rm ion}}{4\pi \alpha_{\rm B} n_{\rm H}^2}\right)^{1/3} \propto M_{\bullet,2}^{1/3}\, n_{\rm H,7}^{-2/3}\, f_{\rm Edd}^{1/3}, \quad 6.$$

where $Q_{\rm ion}$ is the ionizing photon flux, $\alpha_{\rm B}$ is the case-B recombination rate, and $f_{\rm Edd} \equiv L/L_{\rm Edd} \sim O(1)$, providing $R_{\rm ion}/R_{\rm B} \approx f_{\rm Edd}^{1/3}\, n_{\rm H,7}^{-2/3} (M_\bullet/100\ {\rm M_\odot})^{-2/3}$. Thus, for a 100 $M_\odot$ stellar-remnant BH, embedded in a gas cloud with $n_{\rm H} < 10^7\ {\rm cm^{-3}}$, the ionization front expands outside the Bondi radius ($R_{\rm ion} \gtrsim R_{\rm B}$). Accretion becomes intermittent, and the time-averaged rate is orders of magnitude below the original Bondi rate without feedback, remaining in the $\langle \dot{m} \rangle \lesssim 0.5$ regime (Milosavljević et al. 2009a,b; Park & Ricotti 2011; Park & Ricotti 2012; see also Ciotti & Ostriker 2001). Therefore, super-Eddington accretion with a large photon trapping radius cannot be realized in this regime.

However, when the ambient gas is sufficiently dense that the Bondi rate exceeds $\sim 500\ \dot{M}_{\rm Edd}$, 1D RHD simulations find that the inflowing gas structure approaches a steady state without time-dependent oscillations, yielding hyper-Eddington accretion (Inayoshi et al. 2016). 2D RHD simulations have confirmed this conclusion (Takeo et al. 2018) and yielded a more detailed understanding of the accretion morphology (Sugimura et al. 2017a). **Figure 6** shows the 2D distribution of gas density (**Figure 6*a*,*b***) and ionization fractions (**Figure 6*c*,*d***) of accretion flows in this high-density regime. In the early stage (**Figure 6*a*,*c***), ionizing radiation does not reach the Bondi radius owing to effective recombination ($R_{\rm ion} \lesssim R_{\rm B}$). Then, neutral gas accumulated at $R_{\rm ion} < r < R_{\rm B}$ forms a dense shell and collapses onto the center without being prevented by radiative feedback, leading to collapse of the ionized region (**Figure 6*b*,*d***). As a result, steady, isothermal accretion at $\dot{M} \approx \dot{M}_{\rm B} \gtrsim 500\ \dot{M}_{\rm Edd}$ is achieved.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

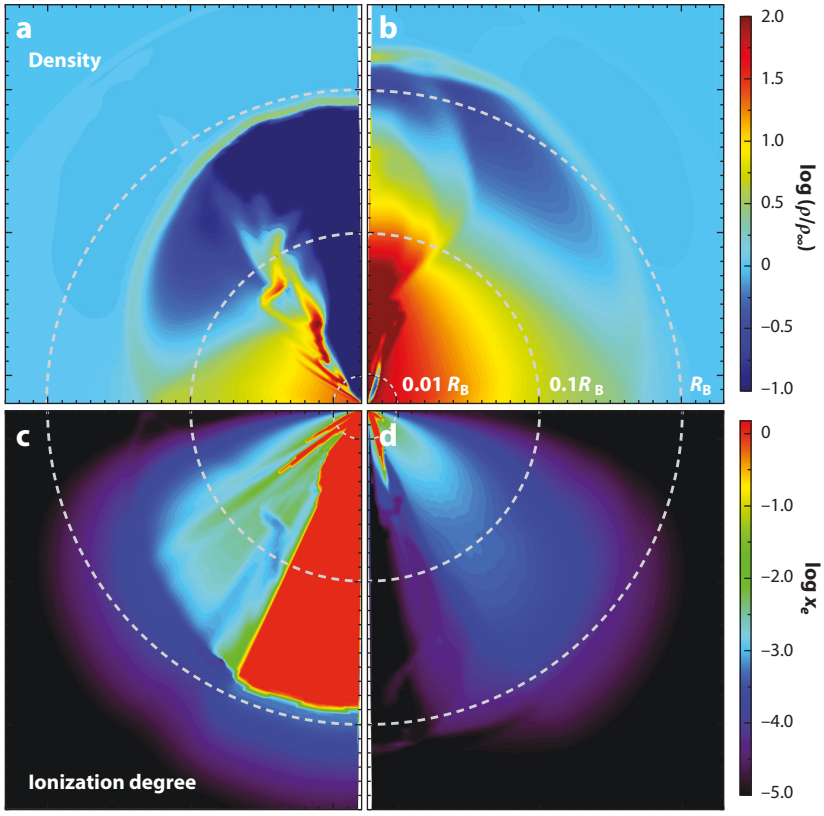

Figure 6

Two-dimensional distributions of (*a*,*b*) gas density and (*c*,*d*) ionization fraction in accretion flows onto a BH in the $\log r$–θ plane. The concentric dashed circles indicate constant fractions of the Bondi radius: $r/R_B = 0.01$, 0.1, and 1.0 from inner to outer circles, respectively. Panels *a* and *c* show the accretion flow during a transition to the hyper-Eddington regime. Even though the ionization front reaches the Bondi radius in the polar regions, rapid gas accretion begins to occur through the equatorial plane, with intense ram pressure. Panels *b* and *d* show the accretion flow after the transition, when the HII region collapses and the entire flow becomes neutral. The accretion rate dramatically increases from time-dependent oscillations at $\langle \dot{M} \rangle \lesssim \dot{M}_{\rm Edd}$ to steady hyper-Eddington accretion at $\dot{M} \sim \dot{M}_B (\sim 500\, \dot{M}_{\rm Edd})$. Data taken from Takeo et al. (2018). Abbreviation: BH, black hole.

This hyper-Eddington accretion solution is stable against radiative and mechanical feedback because photon trapping reduces the emergent luminosity, unless the outward momentum L/c dominates the inward ram pressure of neutral gas (see **Figure 5**). Sakurai et al. (2016a) have found that the inflow rate is not suppressed unless the luminosity emerging at the photosphere is a factor of 10–100 above $L_{\rm Edd}$.[5] **Figure 6** also shows that the anisotropic radiation field has a large impact on the gas distribution near the polar regions but much less in regions near the equatorial plane; i.e., gas flows inward through a disk (see also Sugimura et al. 2017a, 2018). Note that various multidimensional effects (e.g., disk accretion, anisotropy of the radiation field, and Rayleigh-Taylor instabilities), which likely reduce the impact of radiative/mechanical feedback

[5] Radiation heating plays an important role in suppressing inflow gas from the Bondi scale before the transition where $\langle L \rangle \lesssim L_{\rm Edd}$. Therefore, the radiation force onto inflow gas exerted through electron scattering and even bound-free absorption by neutral hydrogen is subdominant until the transition to a hyper-Eddington accretion phase.

on the accreting gas, have been discussed extensively in the context of massive star formation (e.g., Yorke & Sonnhalter 2002, Krumholz et al. 2009). The size of the ionized bubble depends on the spectrum of the radiation emerging from the photosphere, with harder spectra easing the criterion for hyper-Eddington accretion (Takeo et al. 2019). In addition, dust in the accreting gas softens the spectral shape owing to UV attenuation, making the ionized regions smaller (Yajima et al. 2017). Thus, rapid accretion is triggered even for lower BH masses or ambient density unless $Z > 10^{-2}\ Z_\odot$, where super-Eddington accretion is prevented by the radiation pressure of diffuse IR light on dust grains (Toyouchi et al. 2019).

3.1.4. Mechanical feedback. In addition to radiative feedback, BHs accreting at super-Eddington rates can exert negative feedback via winds and jets. In fact, most numerical simulations focusing on the dynamics of BH accretion disks on small scales ($\lesssim 100 R_{\rm Sch}$) find outflows/jets driven by a strong radiation flux and/or a strongly arrested magnetic field. The mechanical feedback associated with BH feeding could play an important role, similar to the low-z AGN population, which is believed to affect large-scale environments, such as star formation on galactic scales (e.g., Fabian 2012, Heckman & Best 2014). Although mechanical feedback has not received much attention in the context of BH growth in high-z protogalaxies, this effect could limit their growth significantly.

Recently, Regan et al. (2019) have investigated the effect of jets launched from an accreting seed BH on gas inflows in an ACH, performing cosmological simulations that resolve the BH gravitational sphere of influence. They found that the momentum injection by jets evacuates a region of approximately ≈0.1 pc surrounding the BH seed, but the jet cannot break out of the halo. Because the impact of the feedback is limited to the vicinity of the BH, the heated and kicked gas will cool and fall back to the center, leading to burst-like accretion episodes as seen in RHD simulations that take into account radiative heating (see Section 3.1.3). As a result, the time-averaged accretion rate over one dynamical timescale (≈100 kyr) at $r \approx R_{\rm B}$ becomes as low as ≈0.2–0.8 $\dot{M}_{\rm Edd}$ (note that Regan et al.'s definition of $\dot{M}_{\rm Edd}$ is 1.6 times higher than ours). This simulation result suggests that a modestly high level of BH accretion at $\approx \dot{M}_{\rm Edd}$ is possible unless the jets impact the gas near the Bondi scale at ≈1 pc. In a longer-duration simulation ($t \gg 100$ kyr), a substantial fraction of the gas could remain within the central 1 pc, and the accumulated mass within the BH influence radius could fall back to the central BH at an even higher rate because ram pressure of inflows would dominate the momentum output of the jets. Although there are still several caveats on the prescriptions for jet mechanical feedback (e.g., mass loading factor and jet energy efficiency), future studies should address these issues and improve our understanding of the early stage of BH growth.

3.2. Stellar-Mass Black Hole Remnants of the First Stars

Having established that super- or hyper-Eddington accretion can occur in principle, we next examine how this may be realized and lead to rapid growth of BHs in the early Universe. A natural and attractive candidate for the initial seed BHs for such rapid accretion is remnants formed in the gravitational collapse of massive Pop III stars, which are the first-generation stars in the Universe (Carr et al. 1984; Omukai & Nishi 1998; Abel et al. 2002; Bromm et al. 2002; Yoshida et al. 2006, 2008; Bromm & Yoshida 2011; see also the review by Greif 2015 and references therein).

In the framework of hierarchical structure formation in the Lambda cold dark matter (ΛCDM) model, the first collapsed baryonic objects are expected to form at $z \gtrsim 20$ in DM minihalos with masses of 10^5–$10^6\ M_\odot$ (Haiman et al. 1996b, Tegmark et al. 1997). The virial temperature of these halos is $T_{\rm vir} =$ few$\times\ 10^2$ K, which is sufficiently high to excite line emission from H_2, formed via gas-phase reactions in the pristine metal-free gas (see more discussion in

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Section 5.2). This emission allows gas to cool and condense in the halo, but because the collapsing gas remains relatively warm ($\gtrsim$100 K), it has a large Jeans mass ($M_J \propto T^{3/2} \sim 10^3$ $M_\odot$), and protostellar cores have a large accretion rate ($\sim 10^{-3}$ $M_\odot$ year^{-1}; see Equation 4). As a result, it has long been thought that the first stars were unusually massive, gaining masses of a few hundred solar masses in their Kelvin–Helmholtz (KH) contraction time of $\sim 10^5$ years. By contrast, a growing Pop III protostar has a high effective surface temperature and emits copious UV radiation. The corresponding ionization and heating of the surrounding gas self-regulates their growth and limits their final masses to $\sim$100 $M_\odot$, with the precise value depending on the ambient temperate and accretion rate (McKee & Tan 2008, Hosokawa et al. 2011, Stacy et al. 2012, Susa et al. 2014). Recent cosmological simulations of Pop III star formation in minihalos have suggested that the IMF indeed tends to be overall top-heavy, with a nearly flat mass distribution in the range of $10 \lesssim M_\star/M_\odot \lesssim 300$ (Hirano et al. 2014, 2015; Stacy et al. 2016), and with the upper end limited by feedback from the protostar's own UV radiation.

Pair-instability supernova explosions (PISNe): extremely luminous supernovae predicted as the end product of low metallicity $\sim$140–260 $M_\odot$ stars

Massive Pop III stars are expected to be short lived ($\sim 10^6$ years) and to promptly endow their parent minihalos at $z \gtrsim 20$ with $M_{\rm seed} \approx$ 10–100 $M_\odot$ seed BHs. Note that very massive nonrotating Pop III stars with $140 \lesssim M_\star/M_\odot \lesssim 260$ may not leave any remnant because of energetic pair-instability supernova explosions (PISNe; Heger & Woosley 2002). Stellar rotation would affect the final fate of massive Pop III stars owing to rotation-induced mixing and extension of the He core. However, the range of zero-age main sequence (ZAMS) mass in which PISNe are predicted to occur is shifted lower only by 10% (Takahashi et al. 2018).

Although the early appearance of such seed BHs is good news, their subsequent rapid growth in minihalos is unlikely. Their Pop III stellar progenitors [as well as any other Pop III star(s) in the same halo] irradiate and blow the ionized gas out of the minihalo, because the gravitational potential well is not sufficiently deep: The sound speed of ionized gas, $\sim$10 km s^{-1}, exceeds the escape velocity, $\sim$1 km s^{-1}, from minihalos (Kitayama et al. 2004, Whalen et al. 2004, Johnson & Bromm 2007). Some remnants are left after energetic SN explosions, which likewise quickly evacuate the gas from the minihalos (Kitayama & Yoshida 2005, Whalen et al. 2008, Ritter et al. 2012). Therefore, the remnant BHs are likely to typically find themselves in exceptionally low-density environments and cannot accrete efficiently. Even if the BHs were to avoid such starvation, and began to accrete, the UV/X-ray radiation associated with this accretion itself would then heat the ambient gas and evacuate the central dense region, self-regulating their growth until their host halos grow much more massive (Alvarez et al. 2009, Jeon et al. 2012, Tanaka et al. 2012). Finally, another obstacle to growth for low-mass BHs is their erratic motion around the central regions (Pfister et al. 2019); as a result, these BHs spend most of the time away from the dense core and accrete inefficiently (Smith et al. 2018).

In principle, mergers of Pop III BHs could also drive their mass growth. However, the merged BHs experience strong recoil kicks with typical velocities of 100 km s^{-1} owing to GW emission, depending on the mass ratio and spin configuration of the merging pair (Campanelli et al. 2007, Herrmann et al. 2007, Koppitz et al. 2007). Because this typical recoil velocity is well above the escape velocity from minihalos, merged BHs will typically be ejected from their parent halos to the IGM, where they cannot accrete at high rates (Haiman 2004). This strongly limits the role of mergers in the early growth of most BHs. However a few ultraearly, rare BHs can avoid this fate by not experiencing mergers until they grow significantly in mass; subsequent mergers will then be at very unequal masses ($\lesssim$1:100), where recoil speeds diminish below a few kilometers per second (leading to a "rich-get-richer" runaway; see Volonteri & Rees 2006, Tanaka & Haiman 2009). In summary, efficient, sustained growth of Pop III remnant BHs likely must wait until these BHs end up in much more massive DM halos, whose potential is deep enough to gravitationally bind enough gas and recoiled BHs.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

3.3. Rapid Growth of Seed Black Holes in High-Redshift Protogalaxies

We here discuss circumstances under which early seed BHs may realize super- or hyper-Eddington accretion, sustained over several orders of magnitude growth in mass. This corresponds to the pathway labelled as hyper-Eddington accretion in **Figure 3**.

The natural place where such rapid BH growth may occur is in the ACHs introduced in Section 1, which are DM halos whose virial temperature is just above the atomic-cooling threshold ($T_{\rm vir} \approx 8{,}000$ K). In a typical ACH, cooling and collapse of the gas is dictated by heavy elements produced in prior episodes of star formation, as well as H_2 molecules formed directly during the collapse. However, in rare cases, prior star formation, as well as H_2 cooling, can be suppressed by several effects, including intense external H_2-dissociating LW irradiation (Haiman et al. 1997, Machacek et al. 2001, Wise & Abel 2007b), streaming motions between DM and baryons (Fialkov et al. 2012, Tanaka & Li 2014), and dynamical heating associated with halo mergers (Yoshida et al. 2003, Wise et al. 2019), or some combination of these effects. The chemistry and thermodynamics of gas in these halos is discussed in detail in Section 5.2 below, in the context of a similar set of requirements for forming a massive $M_\bullet \sim 10^5\ {\rm M}_\odot$ seed BH via an SMS.

Just before the atomic-cooling regime ($T_{\rm vir} \gtrsim 8{,}000$ K), the pristine gas remains essentially adiabatic and settles into a hydrostatic-equilibrium profile in the DM halo's gravitational potential, having a core with $\sim 0.1\ R_{\rm vir}$ and an envelope following $\rho \propto r^{-2}$ (Visbal et al. 2014a), where $R_{\rm vir} \approx 470\ {\rm pc}\ (T_{\rm vir}/8{,}000\ {\rm K})^{1/2}[(1+z)/16]^{-3/2}$ is the halo virial radius. Importantly, the mass of the primordial gas in the core region is as high as $M_{\rm core} \approx 2 \times 10^5\ {\rm M}_\odot\ (T_{\rm vir}/8{,}000\ {\rm K})^{3/2}[(1+z)/16]^{-3/2}$ (Inayoshi et al. 2015), which serves as a rough upper limit on the mass budget available to grow a massive BH. Once the halo crosses the atomic-cooling threshold, atomic-line cooling (primarily Lyα transition) begins to operate, and gravitational contraction of the gas in the core is triggered. The core region develops a density profile as steep as $\rho \propto r^{-2}$, as seen in cosmological simulations of high-z protogalaxies without prior star formation (Wise et al. 2008, Regan et al. 2014a).

The situation envisioned here is that one of the minihalo progenitors of the ACH did manage to form a Pop III star, but its host minihalo remained chemically pristine, because this Pop III star quenched subsequent star formation by its UV radiation (Omukai & Nishi 1999) and then collapsed into a BH without exploding as an SN or ejecting any metals (Heger & Woosley 2002). In fact, this fate may be typical for massive Pop III stars, and therefore several of the minihalo progenitors of the ACH (rather than just one) could have an early star-formation episode, as long as stars in the Pop III stellar IMF that eject metals (low-mass stars and PISNe; see below) are not sampled. In this case, there can be multiple stellar-mass BHs, initially spread spatially in the pristine ACH. However, remnant BHs would quickly decay their orbits owing to dynamical friction on the DM and the gas, depending on their initial orbital properties, and sink to the dense central region. Ryu et al. (2016) simulated the orbital motion of Pop III remnant BHs embedded in gas-rich protogalaxies, taking into account gas drag on the BHs, and found that most initial BH configurations allow one BH (but no more than one) to sink to the center and grow rapidly.

As discussed in Section 3.2, Pop III remnant BHs hardly grow in low-mass minihalos. However, the BH buried in the dense region in the ACH ($\rho \propto r^{-2}$) can be fed at the full Bondi accretion rate, unless BH radiative feedback prevents the inflowing matter. When the ambient matter is self-gravitating, the accretion rate onto the central object is simply given by $\dot{M}_{\rm B} \sim 20\ c_{\rm s}^3/G$ (see also Becerra et al. 2018). In fact, even if H_2 formation is triggered after the halo crosses the atomic-cooling threshold, the accretion rate onto the central BH may remain high, because the Bondi radius is relatively large (~ 0.06 pc; see **Table 3**), and the gas inside this radius is not self-gravitating. As a result, the hyper-Eddington condition, i.e., $\dot{M}_{\rm B} \gtrsim 500\ \dot{M}_{\rm Edd}$, is satisfied until the BH mass reaches $\approx 2 \times 10^5\ {\rm M}_\odot$. When the BH mass exceeds this critical value, the ionized bubble created

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

by the BH expands outside the Bondi radius, where the expansion is further accelerated because of $\frac{d\ln\rho}{d\ln r} < -1.5$ (e.g., Mellema et al. 2006), and the inflowing gas is heated up and would also likely be pushed outward by Lyα photons (Smith et al. 2017). As a result, the rapid hyper-Eddington growth phase of the remnant BH in the ACH is terminated at this mass, independent of the initial seed mass (Inayoshi et al. 2016). By coincidence, this critical mass is comparable with the core mass before the rapid BH growth begins, implying that the bulk of the gas in the core region would be consumed in a hyper-Eddington phase.

In the above works, the collapsing gas had a quasi-spherical geometry. Lupi et al. (2016) considered a similar setup, but with stellar-mass BHs orbiting in a circumnuclear disk, in which clumpy structures form by gravitational instability. They found that the orbiting BHs capture and swallow massive clumps at super-Eddington rates. Combining merger-tree simulations, Pezzulli et al. (2016) have discussed the evolution of seed BHs at high redshifts, including AGN feedback with a simplified model. They concluded that seed BHs with $\sim$100–10^6 $M_\odot$ would likely experience a rapid gas accretion phase in gas-rich protogalaxies. In particular, they find that $\sim$40% (10%) of seeds can grow at hyper-Eddington accretion rates of $\gtrsim 500\,\dot{M}_{\rm Edd}$ at $z = 15-20$ ($z = 10$–15).[6]

In summary, hyper-Eddington accretion can be sustained and quickly produce 10^{5-6} $M_\odot$ BHs by Pop III remnant BHs that find themselves in special ACHs, with chemically pristine gas. Note that these conditions are somewhat less strict than those necessary for the formation of massive seed BHs with $M_\bullet \sim 10^5$ $M_\odot$ via an SMS (discussed in Section 5.2 below), because in the hyper-accreting BH case, some prior star formation occurs in the ACH, and even H_2 cooling may not prevent a brief hyper-Eddington phase once the atomic-cooling threshold is crossed. Nevertheless, pristine ACHs are required, which are very rare at high z. ACHs are more common at lower z, but are more likely to be polluted by metals, yielding a sweet spot for metal-free (or metal-poor) ACHs at $z = 12$–15 (Chon et al. 2016). These redshifts are still sufficiently high to permit further growth to $M_\bullet \approx 10^9$ $M_\odot$ by $z = 6$–7 at the more leisurely Eddington rate.

4. ANGULAR MOMENTUM TRANSPORT

As mentioned in Section 3, in order for infalling gas to maneuver from galactic scales down to the nucleus and accrete onto a compact central object (either a BH or a massive protostar), it needs to shed its large angular momentum. In this section, we discuss the physics of angular momentum transport on multiple scales in the general context of galaxy formation and AGN fueling (Section 4.1), and then specialize to the analogous problem in protogalaxies in the high-z Universe (Section 4.2).

The angular momentum J of an object of mass M and radius R can be specified in terms of a dimensionless spin parameter $\lambda \equiv J/\sqrt{2}Mv_cR$, where $v_c = \sqrt{GM/R}$ is the Keplerian circular velocity. This parameter expresses the level of centrifugal support, with $\lambda = 0$ corresponding to no net rotation, and $\lambda \approx 1$ to full rotational support ($\lambda = 1$ for an isothermal sphere). Due to torques from nearby large-scale structures, DM halos at the time of their virialization acquire a log-normal distribution of λ, with a mean $\langle\lambda\rangle \approx 0.035$, weakly dependent on either halo mass or collapse redshift, from galactic halos in the local Universe (Barnes & Efstathiou 1987, Bullock et al. 2001), down to the halo masses ($\approx 10^6$ $M_\odot$) and up to the redshifts ($z \approx 15$) of interest for the formation of the first massive BHs (Davis & Natarajan 2009).

To illustrate the importance of angular momentum transport for the gas component, it is useful to contrast two scales. First, assuming that the gas shares the halo's specific angular

[6]Note that because Pezzulli et al. (2016) defined the Eddington rate as $\dot{M}_{\rm Edd} \equiv 16L_{\rm Edd}/c^2$, the hyper-Eddington criterion is set at $\dot{M} > 300\,\dot{M}_{\rm Edd}$ in their figure 4.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

momentum (van den Bosch et al. 2002), and that this specific angular momentum is conserved as the gas cools and contracts inside the halo, the centrifugal barrier would halt the collapse at the centrifugal radius, $R_c \approx \lambda R_{vir}$ (Mo et al. 1998). This is a very large radius, $R_c \approx 0.1\ \mathrm{kpc}\ (M_{vir}/10^8\ M_\odot)^{1/3}[(1+z)/11]^{-1} \approx 3 \times 10^{20}$ cm. By comparison, the Schwarzschild radius of a stellar-mass BH of mass $M_\bullet$ is $R_{Sch} = 3 \times 10^7$ cm $(M_\bullet/100\ M_\odot)$, and even the radius of a supermassive protostar (see Section 5.3) with mass $M_\star$ is only $R_\star = 2.6 \times 10^3\ R_\odot\ (M_\star/100\ M_\odot)^{1/2} \approx 2 \times 10^{14}$ cm $(M_\star/100\ M_\odot)^{1/2}$ (Hosokawa et al. 2012).

Centrifugal radius (R_c): the radius below which gas cannot collapse because of angular momentum support if the gas conserves its specific angular momentum

The conclusion is that the gas needs to reach distances $\sim 10^6$ times smaller than the centrifugal barrier in order to be incorporated into a central giant supermassive protostar. The gas also needs to move inward by a further factor of $\sim 10^7$ in order to accrete onto a central stellar-mass BH.

4.1. Angular Momentum Transport in Galaxy Formation

It has indeed long been recognized that efficient angular momentum transport is required to move gas inward by many orders of magnitude in radius, from galactic ($\gtrsim$1 kpc) scales down to the vicinity of a central SMBH, to fuel active galactic nuclei (see, e.g., Shlosman et al. 1990 for a review). In this broader context of galaxy formation, several distinct processes are understood to play important roles, roughly staggered in three distinct spatial scales (which can, however, overlap).

First, on the largest scales, both the gas and the collisionless components (DM and any preexisting stars) develop nonaxisymmetric morphologies. Such nonaxisymmetries are inevitably produced in major mergers (Barnes & Hernquist 1991), but can also develop as a result of perturbations in minor mergers or tidal interactions or even arise in isolated galaxies that already have large self-gravitating disks. Self-gravitating disks or flattened structures are known to be globally unstable to a spontaneous loss of axisymmetry when the ratio of their bulk kinetic energy to potential energy, $T/|W|$, exceeds a critical value (Ostriker & Peebles 1973, Christodoulou et al. 1995). The resulting spiral waves and bar-like structures that develop are known, in turn, to transport angular momentum outward and to facilitate mass inflow (Lynden-Bell & Kalnajs 1972). On galactic scales, the collisionless components (DM and stars) are also important, and misalignments between nonaxisymmetric structures in these components relative to those in the gas provide extra torques that help gas inflow (Shlosman et al. 1989, Barnes & Hernquist 1991).

Second, once the gas has cooled and contracted, it eventually becomes self-gravitating and can develop its own nonaxisymmetric bar-like structures, allowing continued gas inflow (bars-in-bars; Shlosman et al. 1989, 1990). However, the physics of angular momentum transport in this regime is complicated further by the fact that the gas becomes prone to local Toomre instability (Goodman 2003), which could produce fragmentation and efficient star formation (Goodman & Tan 2004). This could consume much of the gas and prohibit the large majority of the gas from crossing this minefield and ever reaching the innermost regions, where the growing BH stabilizes the inner disk (Thompson et al. 2005).

One promising idea is that turbulence, which inevitably develops, facilitates gas inflow. Because the gas needs to cool below the host halo's virial temperature, bulk gas speeds typically exceed the sound speed, and turbulence becomes supersonic. Several hydrodynamical simulations have indeed converged on the following broad picture in this regime (Escala 2007; Mayer et al. 2007, 2010; Levine et al. 2008; Hopkins & Quataert 2010, 2011; Choi et al. 2013, 2015). Large-scale global instabilities generate gas inflow as well as concurrent turbulence down to the smallest resolved spatial scales. This turbulence has been suggested to support the disk against gravitational fragmentation (Levine et al. 2008, Begelman & Shlosman 2009, Choi et al. 2013; although these results may have not yet numerically converged—see further discussion below). As a result, a compact nuclear self-gravitating disk forms, which remains locally stable. The system can reach a

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

quasi-steady state on long timescales, in which global instability drives intermittent bar-like structures. These bar-like structures, as well as turbulence itself, redistribute the angular momentum in the disk on a dynamical timescale and can sustain a large gas inflow rate. In a suite of ∼100 simulations, surveying the parameter space of galaxy properties, Hopkins & Quataert (2010) find a cascade of secondary instabilities, but with a diverse range of nonaxisymmetric morphologies beyond bars that are intermittent, and produce a large (∼1 $M_\odot$ year^{-1}) but correspondingly time-variable accretion rate.

Up to this stage, the presence or absence of a central massive object was immaterial. However, finally, once the gas reaches well inside the sphere of influence of the central BH (if there is one), the disk is no longer unstable to either bar-like modes or local Toomre instability. Hopkins & Quataert (2010) find a new gravitationally driven instability at the boundary of this regime, in the form of a precessing lopsided disk (or one-armed spiral). However, further inside this regime, within $\sim 10^{4-8}$ $R_{\rm Sch}$ of the central BH (depending on BH mass and accretion rate; e.g., Haiman et al. 2009), the disk is gravitationally stable. In this regime, viscosity is understood to be provided by magnetic fields. Even a vanishingly small initial seed field (though amplified by turbulence) is sufficient to generate the magnetorotational instability (MRI; Balbus & Hawley 1991). The MRI and/or related MHD effects can efficiently transport angular momentum in this smallest-scale regime, even at super-Eddington accretion rates (e.g., Jiang et al. 2014).

The major caveat to the above picture is that simulations rely on subgrid prescriptions for cooling, star formation, and feedback on unresolved scales. In particular, simulations typically impose a temperature or entropy floor, which could have a large effect on star formation, angular momentum transport, and the global behavior. These effects can be parameterized and modeled analytically (Hopkins & Quataert 2011).

4.2. Angular Momentum Transport in High-Redshift Protogalaxies

Much of the physics described above also applies in the context of forming massive BHs in the first galaxies. The key differences are that the gas is metal free or metal poor and cools less efficiently; furthermore, gas is unlikely to cool greatly below the halo's virial temperature. As a result, star formation is likely less efficient (at least initially), and the self-gravitating disks that form are likely to be thicker and less prone to instabilities.

Even before discussing angular momentum transport, we note that one way to ease the fueling problem is to start with gas with lower-than-usual angular momentum. Such gas could be found in halos in the low-λ tail of the halo spin distribution (Eisenstein & Loeb 1995) and/or in the low-j tail of the specific angular momentum distribution of gas in individual halos (Koushiappas et al. 2004). A related idea is that in early DM halos, which form at the knots of many filaments of the proto-cosmic web, the gas arrives from many directions along these filaments, resulting in a significant cancelation of the net angular momentum, allowing more efficient initial contraction/inflow (Dubois et al. 2012, Prieto et al. 2015).

Although these effects can help, in order to reach the central massive objects, all three of the above scenarios require significant further outward transfer of angular momentum. The probability distribution of λ is found to be approximately log-normal, but its extreme low spin-tail, where λ is orders of magnitude below the mean, has not been determined from the limited number of halos followed in the above simulations. Nevertheless, it would be unrealistic to appeal to a near-perfect cancelation of large-scale structure torques and then to a conservation of the nearly vanishing specific angular momentum. Further angular momentum transfer is therefore needed and is likely purely gravitational in origin initially (on the largest scales), similar to the picture discussed in the previous subsection. Gas adiabatically condensed in the central regions of

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

pristine ACHs, whose temperature is close to the virial temperature, will remain locally Toomre stable unless they spin exceptionally slowly (Oh & Haiman 2002). However, as long as gas in these early halos can cool and form self-gravitating disks, it can become unstable to global non-axisymmetric modes. These can lead to a redistribution of angular momentum and allow gas inflow to the central region that can ultimately produce a BH (Koushiappas et al. 2004, Lodato & Natarajan 2006), with a range of different BH masses between different halos (Lodato & Natarajan 2007).

More specifically in this context, Begelman et al. (2006) proposed that a multistage cascade of gaseous bars may form and transport angular momentum outward (with gas collapsing down to smaller scales and eventually forming a quasi-star; see Section 5.3 below). This is a follow-up on the bars-in-bars scenario discussed in the previous section. The original proposal consisted of two distinct stages: first a collisionless (stellar) bar driving the gas inward, and then a single gaseous bar forming in a self-gravitating disk (Shlosman et al. 1990). In principle, however, a cascade of several nested bars on increasingly smaller scales could arise, as long as star formation is avoided. Adaptive mesh refinement (AMR) simulations following the central collapse of pristine ACHs have indeed found such a cascade. Wise et al. (2008) identified four nested stages of barlike instabilities, each separated by a factor of ∼100 on successive scales (10^{18}, 10^{16}, 10^{14}, and 10^{11} cm), efficiently driving gas down to the inner region as small as 10^{9} cm. Choi et al. (2013, 2015) have found similar results, confirming the importance of nested gaseous bars.

These simulations also identified supersonic turbulence, which is inevitably produced during the process of virialization (Wise & Abel 2007a) and highlighted its dynamical importance. Note that turbulence both suppresses and stimulates fragmentation. Because turbulence acts as a source of pressure, which counteracts gravity, it tends to stabilize gas against fragmentation, at least on large scales (larger than the size of the turbulent eddies). However, when supersonic turbulent eddies collide, they produce shocks and compress the gas, which promotes fragmentation on small scales (an effect that is absent in the case of thermal pressure). The overall sign of the impact of turbulence depends on whether fragments produced in the latter process can cool and collapse on a timescale shorter than the eddy turnover time (i.e., before they are disrupted by another collision). In the context of star formation in molecular clouds in the local ISM, supersonic turbulence is thought to produce self-gravitating clumps at the high-density tail of the turbulent density distribution, which may explain the shape of the stellar IMF (Padoan & Nordlund 2002). In the high-z protogalaxies, where cooling is inefficient, the net outcome appears to be that turbulence helps stabilize the gas against fragmentation and star formation (Choi et al. 2015). We note, however, that fragmentation may not have been numerically resolved. Indeed, applying the same argument as in the case of the local ISM, at the Mach numbers $M \approx 3$ typical of turbulent inflows in high-z galaxies, one would expect ∼1% of the mass to reside in small self-gravitating fragments, which may be difficult to resolve. Regan et al. (2014a) find that their highest-resolution simulations (with 26 levels of refinement with the open-source code `Enzo`, reaching a resolution of ∼1 AU) point to fragmentation on scales of order ∼100 AU.

We also note that these and similar studies (Regan & Haehnelt 2009, Johnson et al. 2011, Latif et al. 2013) studied ACHs without H_2 chemistry or assuming sufficiently strong LW irradiation so that H_2 has a negligible effect. Though this could be justified in a small subset of ACHs exposed to intense LW radiation (e.g., Shang et al. 2010, Regan et al. 2017) and/or extreme dynamical heating (e.g., Yoshida et al. 2003, Wise et al. 2019), the large majority of early halos will have prior episodes of H_2-cooling-induced fragmentation and star formation in the minihalo stage (see discussion in the next section). Rapid inflow to a central supermassive object is likely inhibited in this case; simulations find a near-Keplerian compact disk fragmenting into dozens of stars growing at subcritical rates (e.g., Greif et al. 2012; see Section 5).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

We end this section by noting that stars are not necessarily only a hindrance for getting gas down to the central BH. In a scenario in which a dense cluster of star surrounds a central seed BH, the stars can help with the angular momentum problem. Alexander & Natarajan (2014) considered the usual Bondi accretion problem, but with angular momentum, and with the inclusion of the acceleration of the BH that would be expected in the presence of a dense star cluster. The gravitational force of the stars results in a jitter in the location of the BH, which, as a result, will see, in its own frame of reference, the angular momentum of some of the infalling gas canceled to zero (or a sufficiently small level for the gas to fall radially inside the Schwarzschild radius). Provided that the gas density is high, so that the Bondi accretion rate is well in excess of the Eddington rate, Alexander & Natarajan (2014) showed that this mechanism can solve the angular momentum problem and permit extended periods of super-Eddington accretion, producing $\approx 10^4$ $M_\odot$ BHs (limited by the need for the star cluster to outweigh the BH).

Stellar-mass/massive black hole seeds: BH seeds $\lesssim$100 $M_\odot$ are defined here as stellar-mass seeds; more massive BHs are referred to as massive seeds

5. THE (INITIAL) MASS FUNCTION OF EARLY BLACK HOLES

In the previous two sections, we reviewed the formation of stellar-mass/massive BH seeds (e.g., Pop III remnants) and general issues regarding their growth via gas accretion, radiative/mechanical feedback, and angular momentum transport of inflowing gas. In this section, we discuss the prompt formation of massive BH seeds, which, for concreteness, we define as any BH heavier than the typical Pop III stellar mass of $\sim$100 $M_\odot$. Motivated largely by the discovery of $\sim 10^9$ $M_\odot$ SMBHs at $z \sim 6$, numerous pathways have been proposed to form such massive seeds (see **Figure 3**). We emphasize that these SMBHs are unusually bright, massive, and very rare objects, having a comoving number density of $\sim$1 Gpc^{-3}(Willott et al. 2010b), but the physics of these pathways is of interest, even if many of these seeds do not actually grow to the masses required to power the high-z quasars.

In this review, we therefore discuss the nature of massive seeds in each scenario and the corresponding IMF of massive BHs over the range $100 \lesssim M_\bullet/M_\odot \lesssim 10^6$ at high redshifts, rather than judge which models can successfully form high-z SMBHs. As emphasized in Section 1.4, investigation of the underlying BH population that does not grow to extreme SMBHs is also crucially important to constrain their BH seeding and growth models with ongoing and future observations (see Section 2) and to better understand the transition between low-z and high-z quasar populations.

In the following subsections, we first review the basic requirements of massive seed formation (Section 5.1) and give a motivation to focus on gravitational collapse of chemically pristine, warm gas in ACHs with virial temperatures of $T_{\rm vir} \gtrsim 8{,}000$ K (see the lower branch of **Figure 3** labelled as atomic-cooling halo). Then, we discuss the possibilities of keeping gas in ACHs warm via suppressed H_2 cooling or enhanced heating (Section 5.2), the resulting emergence of massive seed BHs with $M_\bullet \sim 10^{5-6}$ $M_\odot$ via SMSs (Section 5.3), or with $M_\bullet \sim 10^{3-4}$ $M_\odot$ via runaway mergers in a dense star cluster (Section 5.4), and the subsequent evolution of the population of these BHs in the context of hierarchical galaxy evolution (Section 5.5).

5.1. Prompt Formation of Massive Black Holes

Before going into details, we first describe two general requirements, which are shared by all proposed pathways for massive seed BH formation. First, monolithic collapse of a massive gas cloud is required to form a single massive object, avoiding major episodes of gas fragmentation before the gas reaches very high density. The efficiency of fragmentation depends crucially on the equation of state of the collapsing gas, characterized by the effective heat index $\gamma_{\rm eff} \equiv {\rm d}\ln p/{\rm d}\ln\rho$

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

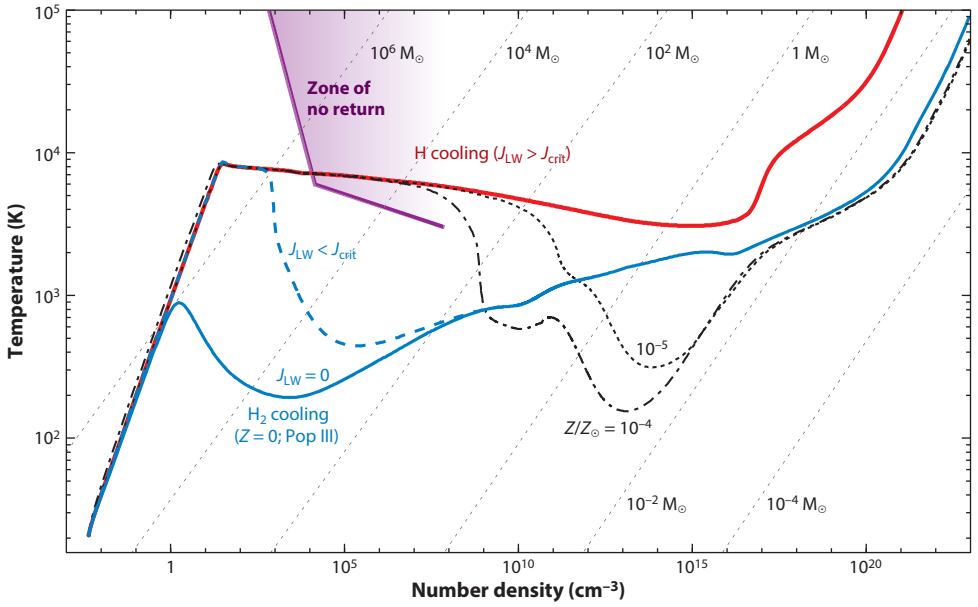

Figure 7

Temperature evolution of a metal-free gas cloud, irradiated by LW radiation with three different intensities (*red curve*, $J_{LW} > J_{crit}$; *blue solid curve*, $J_{LW} = 0$; *blue dashed curve*, $J_{LW} < J_{crit}$), based on a spherical collapse one-zone model assuming a freefall density evolution. With a weak LW intensity ($J_{LW} < J_{crit}$), which dissociates H_2 only at lower densities, gravitational collapse of the cloud is led by H_2 cooling toward higher densities. With a sufficiently high LW intensity ($J_{LW} > J_{crit}$), which keeps H_2 dissociated until the gas enters a dense and hot region (*zone of no return*, demarcated by the *purple line*), the cloud collapses nearly isothermally at $T \approx 5{,}000$ K without rapid cooling phases (*red curve*). Black curves show the evolution of metal- and dust-polluted gas with $Z/Z_\odot = 10^{-4}$ and 10^{-5}, respectively, leading to a rapid temperature drop due to thermal dust emission. In such a rapid cooling phase, the gas is likely to fragment into small clumps, whose masses are approximately the value of the Jeans mass (*dotted diagonal lines*) at the temperature minimum. Abbreviation: LW, Lyman–Werner. Data taken from Omukai et al. (2008).

(Klessen & Glover 2016 and references therein). In rapid cooling phases, where $\gamma_{\rm eff} < 1$, pressure-free and pancake-like collapse of the overdense regions tends to develop a highly flattened sheet-like configuration or filamentary structure. When efficient cooling terminates ($\gamma_{\rm eff} \approx 1$), those filaments refragment into dense cores, each of which tends to collapse in a quasi-spherical way and to not experience further hierarchical fragmentation (Larson 1985, Inutsuka & Miyama 1997). Numerical studies also support that vigorous fragmentation is induced during efficient cooling phases where $\gamma_{\rm eff} \lesssim 1$ (e.g., Clark et al. 2008).

Figure 7 illustrates the n–T phase diagram of a gas cloud collapsing under its self-gravity, obtained in one-zone models under several different conditions. Initially (at the lowest densities), cooling is inefficient, and the gas is heated by compression. For metal-free gas with weak or no LW irradiation (see **Figure 7**) and for slightly metal- and dust-polluted gas, the collapsing gas eventually experiences a rapid temperature drop, caused by cooling via H_2-line or thermal dust continuum emission (see Section 5.2), leading to vigorous fragmentation (e.g., Clark et al. 2008). The fragment mass is approximated by the Jeans mass at the temperature loitering point (where cooling becomes less efficient and $\gamma_{\rm eff} \approx 1$). As indicated in the figure, $M_{\rm J} \sim 10^3$ M$_\odot$ for the H_2-cooling track and $\sim$0.1 M$_\odot$ for the dust-cooling track at this point. Without rapid cooling phases (i.e., due to intense LW irradiation), the thermal evolution is quite different, with no clear single

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

temperature minimum or loitering point. In fact, the collapse is nearly isothermal ($0.9 < \gamma_{\rm eff} < 1.1$) for over ~16 orders of magnitude in density (from ~10 to $\sim 10^{17}\,{\rm cm}^{-3}$). Even in this regime, however, the collapsing central core is unstable against nonspherical perturbations. The collapsing central region elongates slowly with increasing central density, but the amplitude of this distortion may not be large enough to produce fragmentation during the extended isothermal phase (Lai 2000, Hanawa & Matsumoto 2000, Sugimura et al. 2017b).

Second, rapid gas accumulation is required to avoid the formation of a normal massive Pop III star and to instead form an SMS with mass up to $M_\star \sim 10^6\ {\rm M}_\odot$. The critical accretion rate is $\dot{M} \approx 0.01$–$0.1\ {\rm M}_\odot\ {\rm year}^{-1}$ (see Section 5.3), which needs to be sustained at the center of the massive gas cloud collapsing under its self-gravity. The gas accretion rate onto a new-born central protostar is approximately given by $\approx A c_{\rm s}^3/G$ (assuming that the gas was initially in quasi-hydrostatic equilibrium; Larson 1969, Penston 1969; see also Equation 4), with the numerical factor in the range of $A \approx 1$–47 depending on the boundary conditions. 3D simulations typically find $A \approx 20$ (e.g., Inayoshi et al. 2014). The implication is that avoiding normal Pop III star formation requires the gas temperature to remain as high as $\sim 10^4$ K during the collapse phase. This requirement could be satisfied if atomic cooling processes (e.g., Lyα emission) induce gravitational collapse of gas with a suppressed H_2 abundance (**Figure 7**).

As discussed below, these two conditions are fully satisfied only in rare special environments, such as ACHs (*a*) exposed to strong LW irradiation by close neighbors, (*b*) heated through rapid halo mergers, or (*c*) located in regions with an unusually high baryon streaming velocity. This rarity is qualitatively consistent with the fact that SMBHs observed at high redshifts are hosted in rare, very massive galaxies. However, even if the requirements are not achieved perfectly, relatively lower-mass but still massive seeds would form in the IMBH range (10^2–$10^4\ {\rm M}_\odot$). These IMBH seeds could be abundant and have a significant contribution to the overall mass density of the high-z BH population. In addition, the host halos of the high-z quasars formed in highly unrepresentative regions of the Universe, requiring $\sim 5\sigma$ fluctuations on $10^{12-13}\ {\rm M}_\odot$ scales. These special environments may conspire to meet the requirements of massive seed formation and also enable their subsequent growth to the SMBH regime. Indeed, we expect that only a small minority of massive seeds are born in regions that evolve to massive $\sim 10^{12}\ {\rm M}_\odot$ galaxies by $z \approx 6$—the others follow minor branches of the galaxy merger history and remain lower-mass BHs in lower-mass galaxies or in satellite galaxies (Valiante et al. 2016).

In summary, massive seed BHs are expected to result from the near-isothermal collapse of a gravitationally unstable massive gas cloud at a temperature of $T \approx 10^4$ K, which produces a high mass accretion rate onto the central object and avoids major episodes of gas fragmentation. This likely occurs only in rare special environments, as we next discuss.

It is, however, worth keeping in mind a caveat to these requirements. In sufficiently massive galaxies, rapid gas inflow could efficiently fuel a central BH and result in rapid growth, even if the above criteria are not met. Indeed, this must occur for bright quasars at low redshift, which are fueled at rates of $\gtrsim 10\ {\rm M}_\odot\ {\rm year}^{-1}$, despite the gas in the nuclei of their hosts being cold and highly metal-enriched. This contradicts the naive expectation that gas should fragment efficiently in the outer regions of their accretion disks before reaching the innermost regions. High-z galaxies may behave similarly and assemble massive BHs rapidly, despite efficient cooling from metals and/or H_2, once they reach a critical mass. We expect this critical mass to be well above the ACH limit, but this is currently not understood.

5.2. Keeping the Gas Warm: Suppressing H_2 Cooling and Enhancing Heating

It has long been recognized that the key physics governing the formation of the first stars (and BHs) is the abundance of H_2 molecules. This is because collisional excitation of H_2 is the only way

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

for primordial gas to cool radiatively and collapse to high density (Saslaw & Zipoy 1967). Early works have constructed complete gas–phase reaction networks and identified the two possible ways of forming H_2 in primordial gas: via the intermediaries H_2^+ or H^-, of which the latter is relevant in the collapse of high-redshift objects (Peebles & Dicke 1968, Hirasawa 1969, Matsuda et al. 1969). As discussed in Section 3.2 above, the masses of the Pop III stars, formed via H_2 cooling, is tied to the accretion rate of the protostellar core, which is determined primarily by the gas temperature—the cooler the gas, the lower the accretion rate and the stellar mass. Therefore, a key requirement to avoid forming a normal massive Pop III star, and to instead form a much more massive SMS, is the suppression of H_2 formation and cooling. Provided that H_2 is sufficiently suppressed, and the host halo is sufficiently massive (near the atomic-cooling threshold) to raise the gas temperature to several thousand kelvins, the SMS can reach masses of up to $\sim 10^{5–6}\ M_\odot$ (and eventually collapse into a massive seed BH with the same mass; see Section 5.3 below).

Critical Lyman–Werner flux (J_{crit}): the critical value of LW flux above which an ACH collapses nearly isothermally at 10^4 K (due to H_2 photodissociation)

Various H_2-dissociating processes, focusing especially on ACHs, have been investigated by several authors (Omukai 2001, Oh & Haiman 2002, Bromm & Loeb 2003, Schleicher et al. 2010b, Shang et al. 2010, Wolcott-Green et al. 2011, Agarwal et al. 2012, Inayoshi & Omukai 2012, Regan et al. 2014b, Sugimura et al. 2014, Becerra et al. 2015, Latif et al. 2016, Wolcott-Green et al. 2017, Regan & Downes 2018). In particular, three distinct mechanisms could keep the H_2 fraction at low levels: photodissociation of H_2 by LW radiation, photodetachment of H^-, and collisional H_2 dissociation in dense and hot gas. We next discuss these mechanisms, along with other possible ways to keep the primordial gas warm.

5.2.1. Dissociating H_2 by Lyman–Werner radiation. H_2 can be photodissociated by irradiation by soft UV photons in a two-step process. Photons in the ≈11–15-eV range are resonantly absorbed in the Lyman and Werner lines of H_2. These are transitions between the ground and excited electronic states of H_2, analogous to the Lyman series of hydrogen atoms (but split into many rotational and vibrational levels; see below). Roughly ∼10% of the excited $H_2^{(*)}$ decays radiatively into the split state of two H atoms (rather than cascading back to the electronic ground state of H_2), resulting in H_2 dissociation. The significance of this two-step process was first highlighted in the context of the local ISM by Solomon in 1965 (see Field et al. 1966) and subsequently studied by Stecher & Williams (1967).

The earliest generation of stars and remnant BHs emitted LW radiation at redshifts well before cosmic reionization. Because of the long mean free path of photons with energies $h\nu < 13.6$ eV to the IGM, an LW background was built up in the early Universe capable of suppressing H_2 formation in low-mass DM halos (Haiman et al. 1997, Ciardi et al. 2000). H_2 formation in primordial gas occurs mainly through the H^- channel:

$$H + e^- \rightarrow H^- + \gamma, \qquad 7.$$

$$H^- + H \rightarrow H_2 + e^-. \qquad 8.$$

The critical Lyman–Werner flux J_{crit} for suppressing the H_2 abundance follows from balancing the dissociation rate ($\propto J_{LW} \cdot n_H$) with the formation rate ($\propto n_H^2$) and, therefore, depends on the density n_H (linearly, in the optically thin limit). In minihalos, in the absence of H_2 cooling the gas contracts adiabatically,[7] and the relevant density in this case is the maximum value set by the entropy floor of the primordial gas ($n_H \sim 0.1–10\ cm^{-3}$; e.g., Visbal et al. 2014a). Depending on halo mass and redshift, the resulting critical intensity is $J_{crit} \approx (0.01–1)\ J_{21}$, where

[7]Compton cooling on the cosmic microwave background (CMB) is important at high redshift, $z > 10$, but only during the early stages of collapse, because the Compton-cooling time is independent of density.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

$J_{21} \equiv 10^{-21}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} (Haiman et al. 2000, Machacek et al. 2001, Wise & Abel 2007b, O'Shea & Norman 2008, Latif & Khochfar 2019). This is a relatively low value: For reference, the minimum flux required to reionize the Universe, i.e., one ionizing photon per hydrogen atom (neglecting recombinations and the opacity of the IGM), corresponds to a mean background UV flux of $\langle J \rangle \approx 3[(1+z)/11]^3\, J_{21}$. As a result, a large fraction of the earliest minihalos may be sterilized of H_2 and not form any stars (Omukai & Nishi 1999; Ciardi et al. 2000; Haiman et al. 2000; Ricotti et al. 2001, 2002; Mesinger et al. 2006). An important caveat here is that $J_{\rm crit}$ has not been reliably computed in the most massive minihalos just below the ACH. There are, however, hints that in these subatomic halos, $J_{\rm crit}$ rises significantly (to $\gtrsim$100 J_{21}; Regan et al. 2017), especially since H_2 self-shielding can become important (see below). This caveat is important, because massive BH seed formation requires that efficient star formation be avoided all the way to the ACH stage.

The situation is dramatically different for ACHs. As the halo increases its mass and virial temperature to $T_{\rm vir} \approx 8{,}000$ K, Lyα cooling kicks in and gas is able to cool and collapse by collisional excitation of atomic hydrogen, even in the absence of H_2 (Omukai 2001, Oh & Haiman 2002). As the density increases, $J_{\rm crit}$ rises. Additionally, in these halos the column density of H_2 reaches $\sim 10^{14}$ cm^{-2}, at which point the LW lines start to become optically thick (Draine & Bertoldi 1996): H_2 is therefore self-shielded, and the LW flux in the core of the halo is attenuated. The relevant density here is the critical density of H_2 for local thermodynamic equilibrium (LTE), because above this density (*a*) the rovibrational states of H_2 are kept in equilibrium via collisions and radiative cooling becomes ineffective, and (*b*) collisional dissociation from the excited rovibrational levels of H_2 reduces the H_2 fraction (Omukai 2001, Shang et al. 2010, Inayoshi & Omukai 2011). The critical density for the most important transitions is $n_{\rm H} \approx 10^4$ cm^{-3} (Wolcott-Green & Haiman 2019), which is several orders of magnitude higher than the density in minihalos. The value of $J_{\rm crit}$ in the ACHs is therefore correspondingly several orders of magnitude higher.

The thermal evolution of the gas sharply bifurcates, depending on whether the LW intensity is below or above the threshold, as illustrated in **Figure 7**. When $J_{\rm LW} < J_{\rm crit}$, a rapid temperature drop is caused by radiative cooling of self-shielded H_2 after a brief isothermal phase. In this case, the temperature track converges toward the one without LW radiation. As a result of the rapid cooling phase ($\gamma_{\rm eff} < 1$), this gas is expected to fragment into small clumps with $M_{\rm J} \sim 10^3\ M_\odot$ (Regan & Downes 2018, Kulkarni et al. 2019). When $J_{\rm LW} > J_{\rm crit}$, the temperature evolution is qualitatively altered. The nearly isothermal collapse at $T \approx 8{,}000$ K continues until very high density ($\sim 10^{16}$ cm^{-3}) without being affected by H_2 cooling. In this case, fragmentation may be absent or is at least strongly suppressed (Regan & Downes 2018).

There is now a very large literature on the value of $J_{\rm crit}$, which depends sensitively on the detailed calculation of the optically thick H_2–photodissociation rate. In general, this rate must be computed by summing over the rate of resonant absorption into thousands of LW lines, multiplied, in each line, by the probability of eventual decay into the split atomic state. Even in one-zone models, this calculation is challenging: There are 301 rovibrational states of the ground electronic state and over half a million allowed electronic transitions in total. In each LW line, the rate calculation must take into account both the shape of the incident flux (including absorption lines in realistic galaxy spectra overlapping with H_2 lines) and self-shielding by the H_2 line itself (depending on the H_2 column density), as well as possible shielding by the damping wings of neighboring atomic Lyman lines (depending on the H column density). Finally, the H_2 rovibrational levels in the electronic ground state are not all in general in LTE, which significantly affects the effective shielding and the resulting dissociation rate, as well as the radiative cooling. In the 3D case, additional complications arise from bulk motions (which cause Doppler shift on line frequencies),

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

temperature variations, and the basic fact that self-shielding is not a local quantity but rather depends on the direction-dependent column density across the protogalaxy.

Overall, $J_{\rm crit}$ in ACHs has been found in the range of $J_{\rm crit} \sim (10^3$–$10^5)\, J_{21}$. In one-zone models, the most complete calculations, which adopt the most up-to-date chemical network and take into account realistic incident spectral shapes from low-metallicity galaxies, give $J_{\rm crit} \approx$ (1,000–1,400) J_{21} (e.g., Sugimura et al. 2014 and references therein). When combined with calculations of self-shielding (Wolcott-Green et al. 2017) and non-LTE effects (Wolcott-Green & Haiman 2019), the lower end of this range is generally favored (because self-shielding is somewhat weaker when the H_2 is spread over many different rovibrational states). Latif et al. (2015) find the exceptionally low value of 400 J_{21} (with a somewhat different chemistry network). The exact value of $J_{\rm crit}$ indeed depends on the chemical reaction networks and/or reaction rate coefficients adopted in the literature. Uncertainties in the reaction rates translate to a factor of $\approx$2 uncertainty in $J_{\rm crit}$ (Glover 2015a,b).

3D simulations of ACHs find that the collapse dynamics affects the thermal evolution of the gas and can impact $J_{\rm crit}$ significantly. Shang et al. (2010) have noted that turbulent shocks occur at various densities and cause an ∼10–20% scatter in the temperature. If the temperature is lower (high temperature fluctuations are radiated away by Lyα cooling quickly), the collisional dissociation rate is also lower, which requires a higher $J_{\rm crit}$ to compensate. Overall, they found $J_{\rm crit} = (10^4$–$10^5)\, J_{21}$. However, these values are highly sensitive to the treatment of the self-shielding.

Simulations typically adopt a parameter, $f_{\rm shield}$, which depends only on the gas temperature, as well as a local estimate of an effective H_2 column density, to take into account self-shielding in the LW lines. Wolcott-Green et al. (2011) have shown that $J_{\rm crit}$ is reduced by an order of magnitude when a more accurate shielding factor (including excited rovibrational states), as well as a more accurate local column density estimate, is used. Adopting the shielding factors from Wolcott-Green et al. (2011), Latif et al. (2015) have found $J_{\rm crit} \approx (2$–$5) \times 10^4\, J_{21}$. Hartwig et al. (2015) have implemented a new method to capture the gas geometry and velocity field that enables a proper determination of the direction-dependent H_2 self-shielding factor, reducing the critical flux by a factor of two. Overall, the values of $J_{\rm crit}$ in 3D simulations tend to be a factor of a few higher than those in one-zone models when they use the same input spectra and shielding treatment. Importantly, 3D simulations can also include the fact that the LW flux seen from a neighboring galaxy is highly anisotropic (Regan et al. 2016) and have also revealed a strong halo-to-halo variation in $J_{\rm crit}$ by at least a factor of several (Shang et al. 2010, Latif et al. 2014a).

Stellar populations with a significant binary fraction alter the radiation spectra of source galaxies and increase the critical intensity (Agarwal et al. 2017). A unique effect of a significant binary population is X-ray irradiation associated with the LW emitting galaxies. Furthermore, the source galaxies (presumably star-forming galaxies) could be sources of strong cosmic rays, which would have been accelerated in SN remnants originated from the death of massive Pop III stars. Effective ionization by soft X-rays ($\approx$1 keV) and cosmic ray particles enhances the electron fraction and, thus, activates H_2 formation through the electron-catalyzed reactions (Haiman et al. 1996a, Stacy & Bromm 2007, Inayoshi & Omukai 2011, Nakauchi et al. 2014). As a result, the critical flux is boosted by one order of magnitude (Inayoshi & Omukai 2011, Inayoshi & Tanaka 2015, Latif et al. 2015).

The correct knowledge of $J_{\rm crit}$ is crucial in estimating the number of ACHs in which massive seed BHs could form (Dijkstra et al. 2008, 2014; Ahn et al. 2009; Agarwal et al. 2012; Johnson et al. 2013). This is because $J_{\rm crit}$ is substantially higher than the expected level of the LW background at high redshifts well before reionization (Haiman et al. 1997, Wise & Abel 2007b, O'Shea & Norman 2008). Nevertheless, a small fraction of ACHs, which are in an extraordinary overdense region and have bright galaxies nearby, could be irradiated by a sufficiently high flux (Dijkstra

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

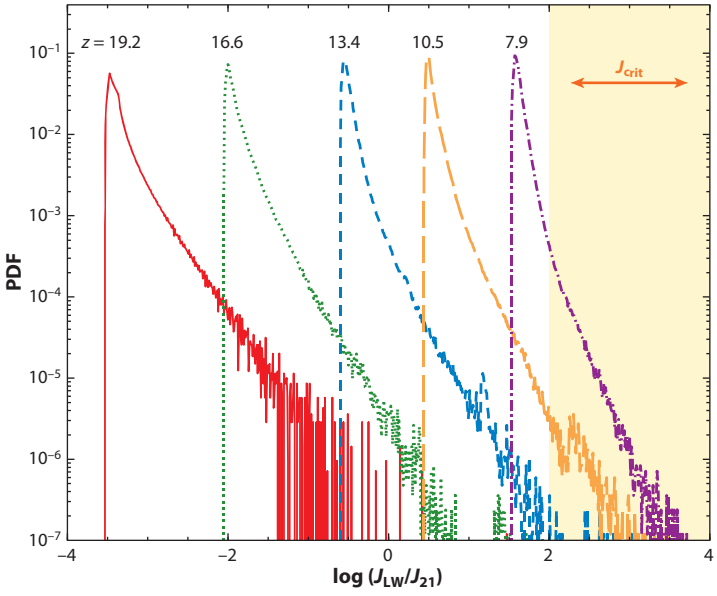

Figure 8

Volume-weighted PDF of the LW radiation background intensity (J_{LW}, in units of J_{21}) at different redshifts in the range of $7.9 < z < 19.2$, obtained in a cosmological simulation. The shaded region on the right marks the expected range of the critical LW intensity J_{crit} for H_2 suppression in ACHs. Because the critical LW flux J_{crit} lies in the far tail of the steep PDF, an exceedingly small but nonzero fraction ($\lesssim 10^{-5}$–10^{-6}) of ACHs at $z \gtrsim 10$ are exposed to sufficiently intense LW radiation and potentially form a massive seed BH. Radiation from subhalos and low-mass minihalos, which were not resolved in the simulation and therefore not included in this figure, can significantly increase this fraction. Figure adapted with permission from Ahn et al. (2009). Abbreviations: ACH, atomic-cooling halo; BH, black hole; LW Lyman–Werner; PDF, probability distribution function.

et al. 2008). The fraction of ACHs that potentially form massive seeds directly reflects the individual properties of a few bright, nearby source galaxies (e.g., stellar mass, star-formation rate, LW luminosity) and their distances from the ACH of interest. The range of required distances is, in fact, quite narrow (Visbal et al. 2014b, Regan et al. 2017). If the neighbor(s) are too far, they must have unrealistically high star-formation efficiencies to produce a sufficiently high LW flux. If they are too close, then the gas in the ACH of interest tends to be either stripped by ram pressure or tidally disrupted by massive source galaxies (Chon et al. 2016) and/or photoevaporated by intense ionizing photons (Johnson et al. 2014, Regan et al. 2016) and polluted by metal winds produced from these neighboring sources (Dijkstra et al. 2014).

Overall, the fraction of ACHs exposed to sufficiently strong LW radiation ($J_{LW} \geq J_{crit}$) sharply decreases with J_{crit}, because the probability distribution function (PDF) of the background flux J_{LW} is very steep (Dijkstra et al. 2008, 2014; Ahn et al. 2009; Agarwal et al. 2012; Johnson et al. 2013; Inayoshi & Tanaka 2015; Chon et al. 2016; Habouzit et al. 2016). **Figure 8** shows an example of this PDF from Ahn et al. (2009). The probability of ACHs having $J_{LW} > J_{crit}$ at $z = 10$ is as small as $\approx 10^{-6}$–10^{-7}. Naively multiplying this by the (comoving) number density of ACHs $n_{ACH} \approx$ few Mpc^{-3} at $z = 10$, the expected number density of massive seed BHs is $n_{seed} \approx$ few Gpc^{-3} $(J_{crit}/10^3)^{-\beta}$, where $\beta \approx 5$ (Inayoshi & Tanaka 2015). However, this neglects any radiation from subhalos, which are unresolved in cosmological simulations of the background. Recent *N*-body (Visbal et al. 2014b) and hydrodynamical (Chon et al. 2016) simulations that resolve subhalos,

as well as lower-mass minihalos, find a significantly flatter flux PDF, and many orders of magnitude higher probability for $J_{LW} > J_{crit}$, due to these extra radiation sources. Many of these halos may fail to collapse, because of tidal disruption and/or ram-pressure stripping, as noted above. However, Visbal et al. (2014b) still find an abundance of massive seed BH-forming halos as high as $\approx 10^{-4}$ Mpc^{-3}.

Visbal et al. (2014b) have further introduced timing as an important aspect, considering only pairs of pristine ACHs that form nearly synchronously (crossing the atomic-cooling threshold within a few million years of each other) and with a small spatial separation (within about a kiloparsec). Regan et al. (2017) have further investigated this synchronized pairs scenario, using cosmological hydrodynamical simulations, and found that a massive seed BH could form only if (*a*) the separation of two halos is within 0.2–0.3 kpc and (*b*) the time for the irradiated ACH to collapse and form a seed BH is within ∼4 Myr, which avoids the deleterious effects of X-ray irradiation, photoevaporation, or metal pollution. However, because these conclusions depend on several uncertain parameters (e.g., IMF, star-formation efficiency, clumping factor of the IGM, metal wind velocity), further studies of this scenario are required to assess how frequently it may ultimately produce massive BHs.

Finally, we note the alternative possibility that the first star(s) within a halo provide the LW radiation that can suppress further fragmentation in the same halo (Susa 2007). Such internal H_2 suppression likely would have to involve significant fine-tuning, because both photoionization feedback and metal enrichment, concurrently with the LW radiation, need to be avoided, whereas a strong LW flux must be present at the ACH stage. Dunn et al. (2018) recently investigated this scenario in hydrodynamical simulations [BHs are treated similar to those in Tremmel et al. (2017), but with LW radiation included] and found massive BHs to be a common outcome, including multiple massive BHs forming in the same halo (above the ACH limit), although their resolution was insufficient to determine the history of ACHs or to follow the formation of a massive seed BH. In order to determine whether massive BHs can form under these circumstances, higher-resolution simulations are required.

5.2.2. Other H_2 suppression mechanisms. We next mention two additional mechanisms that could suppress the H_2 abundance. These effects could reduce the critical LW flux potentially by orders of magnitude, but, as discussed below, they also require special conditions.

5.2.2.1. H^- photodetachment. Low-energy ($\lesssim$11 eV) near-IR and optical photons can not efficiently dissociate H_2 via LW line absorption, owing to the absence of strong LW lines, but can indirectly suppress H_2 formation via H^- photodetachment ($H^- + \gamma \rightarrow H + e^-$). The photon energy threshold for this detachment is $h\nu \geq 0.76$ eV. Several papers computed and quoted a critical flux, J_{crit}, assuming that the incident spectrum is represented by a blackbody shape, with a temperature of either $T_\star \approx 10^5$ K (where the peak frequency is $h\nu_{max} \approx 24$ eV) or $T_\star \sim 10^4$ K ($h\nu_{max} \approx 2.4$ eV). In the latter case, J_{crit} is found to be as low as $30\,J_{21}$ as a result of efficient H^- photodetachment (e.g., Omukai 2001, Bromm & Loeb 2003, Shang et al. 2010). Unfortunately, this low value has caused significant confusion in the literature and created the illusion that Population II (Pop II) galaxies with softer spectra can suppress the H_2 abundance in early galaxies more easily. In reality, the opposite is true: Suppressing the H_2 abundance with softer spectra is more difficult.

Although technically correct, the low J_{crit} values quoted in the steep Wien tail of soft blackbody spectra are misleading. First, because the H_2 suppression is caused by ∼2-eV photons, what matters is the IR intensity and not the LW flux. A spectrum as soft as $T_\star \sim 10^4$ K would be created by low-mass stars, which do not emit much UV radiation. As shown by Wolcott-Green & Haiman (2012), producing the low $J_{LW} = 30\,J_{21}$ with low-mass stars requires a factor of a few

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

more mass in stars than that to achieve the higher $J_{LW}(\approx 10^3)$ with massive Pop III stars. Second, in practice, realistic composite galaxy spectra are not as soft as a $T_\star \sim 10^4$-K black body, unless they are devoid of $\gtrsim 1$ $M_\odot$ stars. Using one-zone models, Wolcott-Green et al. (2017) pointed out that independent of the spectral shape, there is a critical curve in the (k_{LW}, k_{H^-}) plane, where k_{LW} and k_{H^-} are the H_2 dissociation rates by LW and IR photons, which determines whether an illuminated protogalaxy can cool efficiently via H_2. Using population synthesis models for Pop III and Pop II galaxy spectra, the conclusion is, however, that unless the Pop II IMF is even softer than the low-z Salpeter distribution and consists predominantly of ~ 1 $M_\odot$ stars, the direct LW dissociation dominates and H^- detachment plays only a minor role (see also Sugimura et al. 2014, Agarwal & Khochfar 2015).

Zone of no return: region in temperature-density space in which primordial gas cannot cool below ≈3,000 K owing to collisional dissociation of molecular hydrogen

Even if H^- detachment by direct external illumination is not important, the Lyα radiation generated inside the gas cooling in an ACH will be highly trapped, building up a large internal Lyα photon density. In toy models, Johnson & Dijkstra (2017) found that these trapped Lyα photons can detach H^- and reduce the required J_{crit} for the external UV flux. Recent 3D simulations confirm this conclusion and find reductions of ∼20% in J_{crit} (Wolcott-Green et al. 2020).

5.2.2.2. Collisional H_2 dissociation. Alternatively, H_2 collisional dissociation ($H + H_2 \rightarrow 3H$) potentially plays an important role in the formation of massive seed BHs in dense ($n_H \gtrsim 10^4$ cm^{-3}) and hot ($T \approx 8{,}000$ K) shocked regions (Inayoshi & Omukai 2012). Such shocked regions may result from colliding inflows at the centers of protogalaxies in their assembly and/or via violent collisions of galaxies themselves (Mayer et al. 2010, Inayoshi et al. 2015, Mayer et al. 2015). In the primordial case, if the temperature and density of the metal-free postshock gas are high enough for H_2 rovibrational levels to reach LTE, the gas never cools down below ≈3,000 K because of the lack of H_2 cooling due to collisional dissociation. This "zone of no return" is marked in **Figure 7**. The shocked layer fragments and forms a massive cloud with $\sim 10^5$ $M_\odot$, which collapses near-isothermally via atomic cooling as in the LW-aided scenario. Fernandez et al. (2014) investigated this shock-aided scenario with 3D cosmological simulations and confirmed the basic idea of the zone of no return. However, they found that for ACHs with virial temperatures near the atomic-cooling threshold (≈8,000 K), cold gas flows accrete into the halo but experience shocks before reaching the central region at densities that are too low. This, by contrast, may happen because the radiative cooling timescale is not short enough for hot gas heated by virial shocks to collapse well inside the halo [see Visbal et al. (2014a) for sub-ACHs, and Birnboim & Dekel (2003) and Dekel & Birnboim (2006) for more massive DM halos with $\sim 10^{12}$ $M_\odot$ at lower redshifts]. In more massive halos, with $T_{vir} > 10^4$ K, where the shock-dissipated energy is more quickly carried away by radiative cooling (the cooling rate is a steep function of temperature, $\Lambda \propto T^\beta$ with $\beta \sim 8$ at 8,000 K $< T < 2 \times 10^4$ K, and the cooling timescale of primordial gas has the minimum value of $T \simeq 2 \times 10^4$ K), the colliding inflows may still produce the required high-density, high-temperature shocked gas in the core.

5.2.3. Baryonic streaming motions. An alternative way to avoid star formation in early galaxies is provided by the large relative velocities of baryons with respect to DM, which develop in the wake of cosmological recombination at $z \approx 1{,}100$ (Tseliakhovich & Hirata 2010). These streaming motions are coherent on about megaparsec scales with typical (root-mean-square) magnitudes of ≈30 km s^{-1} $[(1 + z)/1{,}100]$ and can delay the collapse of gas into early DM minihalos (Greif et al. 2011, Stacy et al. 2011, Fialkov et al. 2012). The resulting lack of star formation can keep the halo gas pristine when it finally collapses into more massive halos near the atomic-cooling threshold of $T_{vir} \approx 8{,}000$ K (Tanaka & Li 2014, Schauer et al. 2017). Furthermore, the value of the streaming velocity has a Gaussian distribution, and therefore in rare, high-velocity patches

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

of the Universe, the onset of gas collapse is further delayed until the DM halos become as massive as $\approx 10^8$ $M_\odot$ ($T_{vir} \approx 2 \times 10^4$ K at $15 < z < 20$), which is a factor of ~10–30 above the atomic-cooling threshold (Hirano et al. 2017). In this regime, dynamical effects due to frequent mergers of gaseous halos violently disturb the gaseous cores of the interacting galaxies and further prevent star formation (Hirano et al. 2018). Inayoshi et al. (2018) used Monte Carlo merger trees to simulate the assembly history of DM halos with streaming velocities at twice the root-mean-square value. The fraction and absolute number density of pristine halos with $T_{vir} \approx 2 \times 10^4$ K are estimated as $\sim 3 \times 10^{-5}$ and $\approx 10^{-5}$–10^{-4} Mpc^{-3} at $15 < z < 20$. In such massive halos, well above the atomic-cooling threshold, efficient Lyα cooling could drive cold, pristine streams penetrating deep inside the halo and directly feeding a dense central galactic disk, where massive seed BHs might form from the dense and warm shocked gas, surrounded by a massive disk of Pop III stars.

5.2.4. Rapid galaxy assembly. Finally, a yet different way to avoid star formation in early galaxies is through unusually rapid galaxy assembly (note that the existence of streaming motion is not required here, though high streaming velocities would likely trigger violent mergers of gaseous halos as discussed in Section 5.2.3). If mergers are sufficiently frequent, they may continuously interrupt H_2 cooling and heat the gas back up to the virial temperature (Fernandez et al. 2014) and/or counteract any cooling via enhanced compressional heating (Yoshida et al. 2003, Chon et al. 2016, Wise et al. 2019). In particular, Chon et al. (2016) found that frequent, relatively minor mergers of ACHs generally decrease the gas density at the core via dynamical heating and prevent its gravitational collapse. Collapse of the cloud is induced only in rare cases of frequent major-merger events in which the core mass is boosted by a factor of $\gtrsim 10$ within one dynamical timescale. Recent numerical simulations by Hirano et al. (2017) and Wise et al. (2019) have indeed suggested that the combination of dynamical heating and a weak LW radiation background with $J_{LW} \ll J_{crit}$ increases the gas temperature on ~10 pc scales, enhancing the mass inflow rate toward the center. Mayer et al. (2015) also proposed that short-cutting their prior history, shock heating during Milky-Way-size galaxy mergers at $z \approx 6$ plays a major role in balancing cooling, keeping the temperature of the core at $T \approx 8{,}000$ K, and producing supersonic infall at rates as high as $>10^4$ $M_\odot$ $year^{-1}$, even if the gas is heavily metal polluted. However, such rapid mass accretion likely leads to the formation of a gravitationally unstable nuclear disk, and accretion then might be quenched by efficient fragmentation (Ferrara et al. 2013).

5.3. Massive Black Holes Via a Supermassive Star

Assuming that H_2 suppression, extra heating, streaming motions, or some combination of these kept the gas warm in an ACH, the next question is whether this gas fragments into clumps or collapses monolithically into one single object. Here, we discuss the formation of the protostar, its growth into an SMS, and its eventual collapse into a massive BH.

5.3.1. Prestellar collapse: the birth of a protostar. 3D hydrodynamical simulations have found that an H_2-free cloud exposed to intense LW radiation could avoid fragmentation and continue to collapse monotonically (Bromm & Loeb 2003, Wise et al. 2008, Regan & Haehnelt 2009, Shang et al. 2010, Choi et al. 2013, Latif et al. 2013). However, the issue is not settled, because most of these cosmological simulations have utilized simplifying assumptions in studying the fragmentation process (such as turning off H_2 cooling by hand or adopting an optically thin treatment of Lyα cooling) and had limited spatial resolution [astronomical-unit-resolution runs by Regan et al. (2014a) point to possible fragmentation on scales of ~100 AU]. Inayoshi et al. (2014) have studied gravitational collapse of a warm primordial gas cloud with $T \sim 8{,}000$ K up to densities

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

high enough for the gas to become optically thick and form a protostar ($n_H \gtrsim 10^{16}$ cm^{-3} and $r \lesssim$ 0.1 AU) using a 3D simulation that includes all the relevant cooling processes of both H_2 and H but adopting idealized, noncosmological initial conditions with a weakly turbulent field. Van Borm et al. (2014) have performed similar simulations with somewhat different initial conditions and simplified treatments of the chemical reaction network and radiative cooling.

There are several crucial findings: The central core collapses almost isothermally ($T \approx$ 5,000–8,000 K) until $\sim 10^{16}$ cm^{-3}, forms one single object without major episodes of fragmentation, and accretes onto the central protostar at a high rate of $\sim$1 $M_\odot$ year^{-1} ($\approx 20\, c_s^3/G$).

Figure 9 presents thermal evolution of a pristine, massive collapsing gas cloud in which H_2 is initially dissociated by intense LW radiation, obtained from a 3D hydrodynamical simulation. In the low-density regime ($n_H \lesssim 10^4$ cm^{-3}), the cooling is mainly via Lyα emission, as assumed in most previous work. In fact, Lyα cooling becomes less efficient and continuum cooling via two-photon emission leads to further collapse until $n_H \sim 10^8$ cm^{-3}. At higher density, the dominant cooling process shifts to the H^- free-bound emission ($H + e^- \rightarrow H^- + \gamma$). For $n_H > 10^{15}$ cm^{-3}, photons from the H^- free-bound emission are self-absorbed, as well as Rayleigh scattered by neutral H. The gas collapse proceeds further by H^- free–free emission ($H + e^- \rightarrow H + e^- + \gamma$) until $\sim 10^{16}$ cm^{-3} with the temperature decreasing gradually to $\sim$3,000 K. Finally,

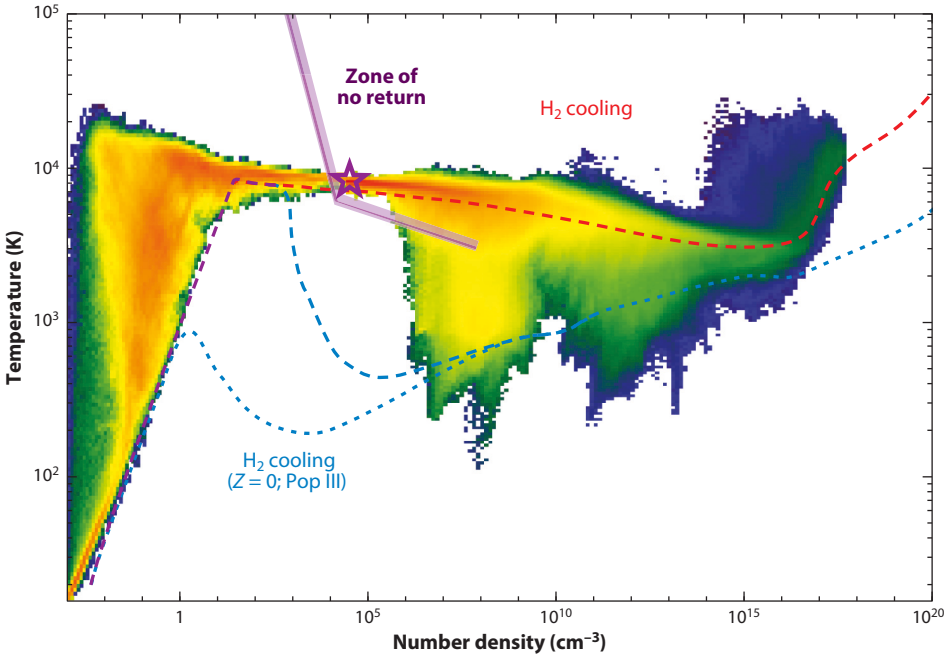

Figure 9

3D simulation result showing the density-temperature phase diagram of a massive collapsing cloud in an ACH. If the gas in the core can avoid H_2 cooling by some mechanisms (e.g., strong LW irradiation in this case) until it enters the zone of no return (*purple shaded line*), then subsequent H_2 cooling can be naturally averted. Most of the collapsing gas resides in the hot component ($T \approx$ 5,000 K), which ultimately forms a rapidly accreting protostar at the center. The overall behavior is essentially consistent with the one-zone models shown in **Figure 7** (note that if H_2 cooling is efficient, then the simulation results are also consistent with the H_2-cooling track shown in **Figure 7**). Data taken from Fernandez et al. (2014) at $n_H < 10^4$ cm^{-3}, and Inayoshi et al. (2014) at $n_H > 10^4$ cm^{-3}. The two results are connected at the boundary of the zone of no return (*star symbol*). Abbreviations: ACH, atomic-cooling halo; LW, Lyman–Werner.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

at this stage, the cloud becomes completely opaque to all continuum emission and forms an adiabatic core, i.e., a central protostar, with a mass of $\sim$0.2 $M_\odot$.

At the beginning of the cloud collapse ($n_H \gtrsim 10^4$ cm^{-3} in the zone of no return), the H_2 fraction is at its equilibrium value ($x_{H_2} \approx 10^{-8}$, which is almost independent of density), balancing formation through H^- and collisional dissociation. This H_2 fraction is too small to cool the gas via line emission. As the density reaches $\sim 10^{11}$ cm^{-3}, the H_2 fraction jumps up to $x_{H_2} \sim 0.1$ by the three-body reaction ($3H \rightarrow H_2 + H$) in the inner region ($\lesssim 10^3$ AU). However, neither the H_2 line nor collision-induced emission (CIE) cooling plays a significant role in the thermal evolution: H_2 lines are optically thick at $n_H > 10^{14}$ cm^{-3}, and other continuum cooling is more important than the H_2 CIE cooling. Combined with adiabatic cooling due to turbulent expansion, radiative cooling related to H_2 induces thermal instability, producing cold gas with $T < 10^3$ K over the wide density range of 10^6 cm^{-3} $< n_H < 10^{13}$ cm^{-3}, where the temperature evolution deviates significantly from the one-zone model results, as shown in **Figure 9**. Because the cold components are not massive enough to be gravitationally bound, the evolution of the central collapsing region is ultimately not affected (see Inayoshi et al. 2014, their figures 3 and 4).

Once the dense core becomes optically thick, the effects of the trapped radiation become important. Luo et al. (2018) performed AMR simulations with `Enzo`, in which the effects of the radiation were included via the FLD approximation. These simulations exclude H_2 and adopt tabulated opacities for equilibrium abundances (i.e., they do not follow the nonequilibrium H^- fraction), but they reach resolutions of 0.01–0.1 AU and resolve the complex shape and time dependence of the photosphere. They find qualitatively similar results to the above in their adiabatic control run. By contrast, when radiation via FLD is included, they find near-Eddington luminosities in the core, with the radiation escaping in directions of the steepest density and temperature gradients. This leads to intermittent outflows, originating near the photosphere, while the core inside the photosphere is quasi-spherical (rather than disky) and has very little rotation.

The prestellar collapse of H_2-suppressed gas has also been investigated in high-resolution cosmological simulations (Regan et al. 2014a, Becerra et al. 2015, Latif et al. 2016), which included H_2 chemistry and cooling and have converged on similar results, as illustrated in **Figure 10**. The early stages of collapse essentially agree with previous noncosmological simulations. This is because the dynamics of the collapsing gas obeys a self-similar solution, where the initial and boundary conditions have been forgotten (Larson 2003). Namely, the density profile consists of the central core and accreting envelope with the $\rho \propto r^{-2}$ law (Larson 1969), and the rotational velocity is as large as half the Keplerian velocity (Narita et al. 1984, Saigo & Hanawa 1998), which agrees with the expected universal value (Abel et al. 2002, Yoshida et al. 2008). This stage is illustrated in **Figure 10*a–d***, down to $\lesssim$1,000 AU, as well as in **Figure 10*e***. Subsequently, when the bulk of the envelope mass is accreting onto the central region and protostar, gases are reminded of their initial and boundary conditions. The infalling matter forms a compact accretion disk, surrounding a central embryonic protostar, but first still without undergoing a major episode of fragmentation (see **Figure 10*c,d***). Because of the high accretion rate (>0.1 $M_\odot$ year^{-1}), the disk, however, soon becomes massive enough to be unstable under its self-gravity and is likely to fragment into smaller clumps even when H_2 cooling does not play an important role (see **Figure 10*f–h***).

Numerical limitations have precluded following the subsequent evolution for longer than $\sim$200 years after this stage (Regan et al. 2014a). Inayoshi & Haiman (2014) discussed the evolution of clumps in the disk with an analytical model, taking into account the growth of clumps via accretion and inward migration. The clumps can rapidly migrate inward on a timescale of $\sim 10^{4–5}$ year, which is shorter than the internal KH timescale in the clumps. Therefore, most of the clumps can merge with the central protostar before forming massive stars. The clumpy structure of the disk

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

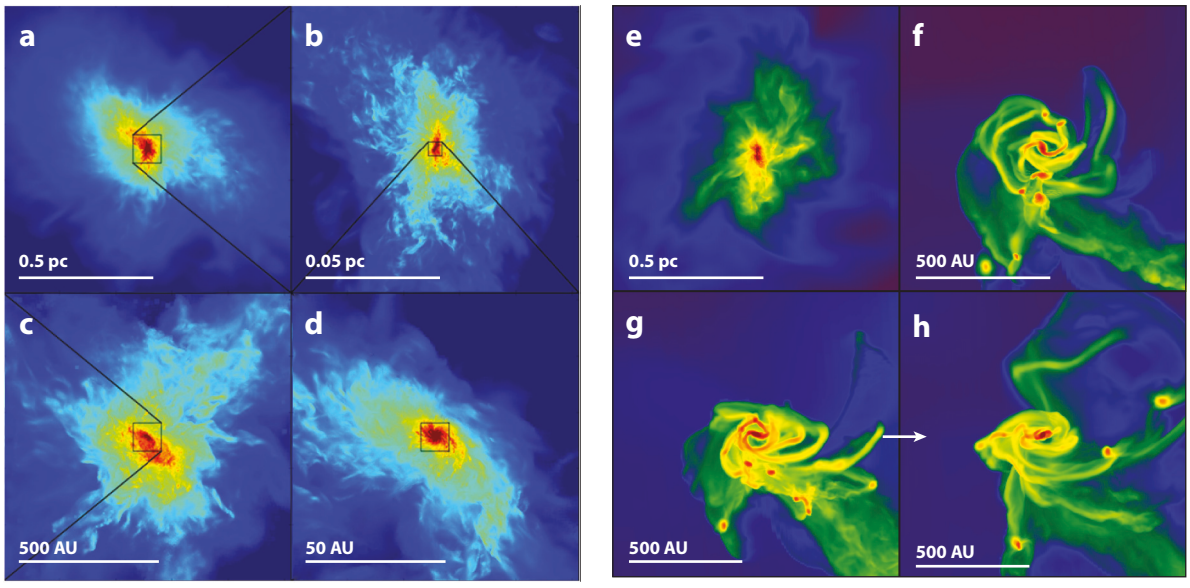

Figure 10

Distribution of gas density in an ACH in which H_2 cooling is suppressed by strong LW radiation with $J_{LW} \gtrsim J_{crit}$. The results in panels *a–d* and *e–f* are from two different simulations, which overall find similar behavior. The central collapsing region does not undergo a major episode of fragmentation despite the complex flow structure caused by turbulence (panels *a*,*b*, down to a few 1,000 AU scales, and panel *e*). However, eventually a disk forms (panels *c*,*d* and *f–h*), which accretes at a high rate, becomes gravitationally unstable, and soon fragments into clumps (best seen in the right-hand panels). The fragments are found to migrate quickly toward the center and are expected to coagulate to form a single supermassive star. Panels *a–d* adapted from figure 1 of Becerra et al. (2015), and panels *e–h* adapted from figures 7 and 11 of Regan et al. (2014a). Abbreviations: ACH, atomic cooling halo; LW, Lyman–Werner.

at a high accretion rate provides episodic burst-like accretion, affecting the protostellar evolution (Sakurai et al. 2016b).

As noted above, an important caveat is that the trapped radiation can become dynamically important inside the optically thick core and can have a strong impact on the nature of the central object (i.e., its angular momentum, shape, density, and accretion rate). Ardaneh et al. (2018) performed cosmological versions of the AMR simulations by Luo et al. (2018) and found qualitatively similar results. In particular, the core structure is irregular and strongly shaped by recurrent outflows driven by both radiation and thermal pressure. Such a rapidly accreting and, thus, spinning-up protostar would also lead to nonaxisymmetric deformation to a bar-like shape that enables efficient angular momentum transfer to the surrounding medium (Lin et al. 2011). These outflows mix with the inflow and are ultimately trapped, but they help outward transfer of angular momentum and result in a rapidly accreting, quasi-spherical central object without significant rotation.

5.3.2. Growth of a rapidly accreting protostar. What is the fate of the protostar surrounded by unlimited amounts of gas? Theoretically, such a rapidly accreting protostar is expected to evolve into an SMS, for which the entropy input by rapid accretion and energy generation by nuclear burning support the entire stellar structure, before it ultimately collapses to a BH.

Figure 11 shows the evolution of the radii of protostars accreting at different rates, based on spherical stellar-evolution models by Hosokawa et al. (2012, 2013). In the ordinary Pop III star case, where $\dot{M} \approx 10^{-3}\ M_\odot\ \text{year}^{-1}$ is set by H_2 cooling, the protostar initially expands as it

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

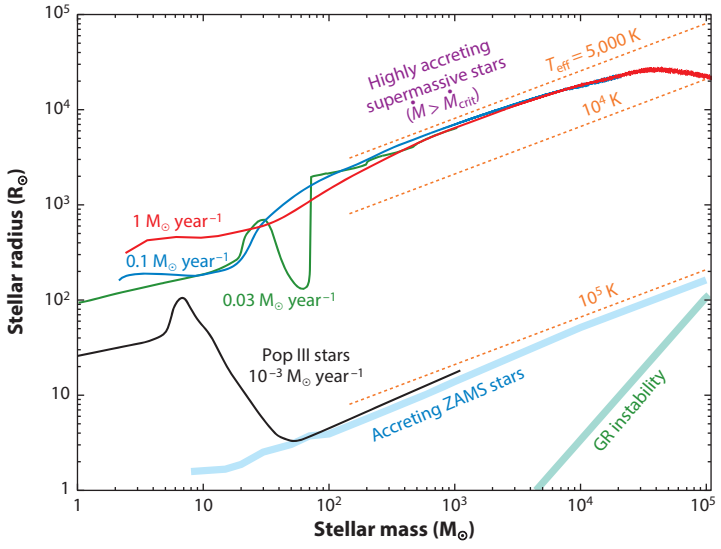

Figure 11

Evolution of the protostellar radius for several different accretion rates in the range of $10^{-3} \leq \dot{M}_\star/\mathrm{M}_\odot$ year$^{-1} \leq 1.0$. For the lowest accretion rate (ordinary massive Pop III stars; *black curve*), the stellar structure contracts and settles down to that of a ZAMS star with a high effective temperature ($\approx 10^5$ K). For higher accretion rates ($\dot{M}_\star > \dot{M}_{\rm crit}$), the protostar continues to expand until its mass reaches $M_\star \sim 10^5$ $\mathrm{M}_\odot$, above which the core region (which contains most of the mass) enters the GR instability regime. The bloated envelope has a low temperature of $T_\star \approx 5{,}000$ K, for which UV stellar feedback does not halt gas accretion (the effective temperature indicated by *orange lines* is estimated assuming $L_\star = L_{\rm Edd}$, which is a good assumption for massive stars with $M_\star > 10^2$ $\mathrm{M}_\odot$). Abbreviations: GR, general relativity; Pop III, Population III; ZAMS, zero-age main sequence. Data taken with permission from Hosokawa et al. (2012, 2013).

gains mass, due to adiabatic heat input by accretion. At $M_\star \sim 10$ $\mathrm{M}_\odot$, it begins to contract by cooling via radiative diffusion (the so-called KH contraction phase) until nuclear ignition occurs at the center (Stahler et al. 1986; Omukai & Palla 2001, 2003). In contrast, at higher accretion rates of $\dot{M} \gtrsim 0.03$ $\mathrm{M}_\odot$ year^{-1}, the growing protostar continues to expand without any KH contraction (Hosokawa et al. 2012). In fact, the interior material contracts and increases the central temperature to the onset of nuclear burning, whereas the outermost layers significantly swell up, resembling a red giant star. This is because the outermost envelope absorbs a part of the outward heat flux and gains energy from the accreted material.

Importantly, the effective temperature of the bloated atmosphere is almost constant at $T_\star \sim 5{,}000$ K, regardless of the initial mass, due to the strong temperature dependence of H^- bound-free absorption opacity (Hayashi 1961). As a result, the ionizing flux is reduced by $\gtrsim 8$ orders of magnitude compared to a ZAMS Pop III star with the same mass. Thus, the UV feedback that could limit the stellar masses at lower accretion rates to at most a few 100 $\mathrm{M}_\odot$ (Hirano et al. 2014; see also McKee & Tan 2008, Hosokawa et al. 2011) never operates until the mass reaches $M_\star \gtrsim 10^5$ $\mathrm{M}_\odot$, where general relativity (GR) instability induces collapse (Hosokawa et al. 2013). The accreting protostar with its bloated envelope is pulsationally unstable, similar to red giants, due to the κ mechanism excited in the He^+ ionization layer in the envelope (Inayoshi et al. 2013). However, the mass-loss rate is significantly lower than the mass accretion rate. In summary, the growth of an accreting SMS is not prevented by either UV feedback or pulsational instability.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Gas accretion, in reality, likely proceeds through an accretion disk. Because the mass inflow rate onto the disk from the parent cloud is high, the disk becomes gravitationally unstable and is likely to feed the central protostar via episodic mass accretion (Regan et al. 2014a, Inayoshi & Haiman 2014, Becerra et al. 2015, Latif & Schleicher 2015, Latif et al. 2016; see also **Figure 10**). If the intervals of episodic accretion due to clump migration through the disk are sufficiently long, the bloated envelope of the SMS would contract and emit strong UV radiation. Sakurai et al. (2016b) have performed simulations of a circumstellar disk and the structure of the central SMS self-consistently by reflecting burst-like accretion episodes. Because the typical interval of episodic accretion is shorter than the local KH-contraction timescale in the protostellar surface layer (where gas opacity is high due to H^- bound-free absorption, and the KH-contraction timescale is an order of magnitude longer than averaged over the entire star), the stellar UV feedback never inhibits gas accretion.

The bifurcation in the evolution is determined by the accretion rate. The interior stellar structure within an inflating envelope was discussed by Omukai & Palla (2001, 2003), who found a critical rate of $\dot{M} \approx 4 \times 10^{-3}\ \mathrm{M_\odot\ year^{-1}}$, above which the total luminosity (interior + accretion) during the KH contraction marginally exceeds the Eddington luminosity. Hosokawa et al. (2012, 2013) found that the inflated structure appears and stably exists only at higher accretion rates of $\gtrsim 3 \times 10^{-2}\ \mathrm{M_\odot\ year^{-1}}$. Recently, Haemmerlé et al. (2018) found that at somewhat lower $\dot{M} \approx 10^{-2}\ \mathrm{M_\odot\ year^{-1}}$, the stellar envelope is still bloated by rapid entropy input at $M_\star \gtrsim 10^3\ \mathrm{M_\odot}$ unless UV feedback and pulsational instability prevent mass accretion. These accretion rates may, in fact, be realized in some of the massive minihalos and/or in ACHs in which H_2 cooling is not fully suppressed and the temperature is as low as a few 10^3 K, potentially leading to many IMBHs with $10^2 \lesssim M_\bullet/\mathrm{M_\odot} \lesssim 10^5$, as remnants of SMSs with intermediate accretion rates in these halos.

5.3.3. The final fates of growing supermassive stars. According to the classical argument (Chandrasekhar 1964, Zel'dovich & Novikov 1971, Shapiro & Teukolsky 1983), an SMS exceeding a critical mass of $M_{\rm GR}$ is thought to directly collapse to massive BHs via a GR instability.[8] The critical mass is on the order of $\sim 10^5$–$10^6\ \mathrm{M_\odot}$, depending on the detailed properties of the stellar rotation and radial structure. Shibata & Shapiro (2002) investigated the gravitational collapse of a rotating SMS in full GR simulations and found that most of the stellar mass is eventually swallowed by the newly born BH, ejecting only ∼10% of the mass. Note that even if the star initially has fast differential rotation, the angular momentum would be quickly transported by turbulent viscosity driven by MRI, but a higher fraction of the initial rest mass of the star forms a disk instead of directly collapsing into the BH (Sun et al. 2018). If the star is rotating sufficiently fast at the beginning of gravitational collapse, the SMS collapses and turns into an intermediate stage in which a close binary BH forms (Reisswig et al. 2013). Some authors proposed a different picture, in which only the central part of the SMS collapses to form a smaller ∼100 $\mathrm{M_\odot}$ BH and the outer envelope is still inflated by energy input from the accreting BH (quasi-star; Begelman et al. 2006, 2008). However, these results are based on a simplified treatment of the equation of state; namely, a fully convective star with a homogeneous entropy distribution is assumed. Recent stellar evolution calculations suggest that rapid mass accretion onto SMSs drastically changes their stellar structure, causing significant bloating with a positive entropy gradient (Hosokawa et al. 2013, Haemmerlé

[8] In the interior of a very massive star, radiation pressure dominates gas pressure, leading to the adiabatic index $\Gamma_{\rm ad} \approx 4/3 + \beta/6$, where $\beta \equiv P_{\rm gas}/(P_{\rm rad} + P_{\rm gas}) \approx 0.027\ (M_\star/10^5\ \mathrm{M_\odot})^{-1/2}$. By contrast, in the relativistic regime, the critical index against a small radial perturbation is $\Gamma_{\rm crit} \approx 4/3 + 1.12\ (R_{\rm Sch}/R)$, where the second term comes from the GR effect. Note that in the Newtonian limit for less massive stars, $\Gamma_{\rm ad}(=5/3)$ is larger than $\Gamma_{\rm crit}(=4/3)$, where the stellar structure is stable. Refer to **Figure 11** where the critical conditions for the GR stability are shown ($\Gamma_{\rm crit} > \Gamma_{\rm ad}$; Fricke 1973).

et al. 2018). As the homogeneous convective core gradually extends after the ignition of hydrogen burning, not the entire star but only its inner core might collapse due to the GR instability.

Although runaway nuclear fusion might cause a very energetic SN explosion for an SMS with high metallicities, this is less likely at zero or low metallicity (Montero et al. 2012). Umeda et al. (2016) extended the stellar structure calculations to the onset of the GR instability and found the critical mass $M_{\rm GR}$ increasing monotonically with stellar accretion rate (see also Woods et al. 2017): (*a*) at $\dot{M} < 0.1$ $M_\odot$ year^{-1}, the stellar collapse begins to occur when the nuclear fuel is exhausted, (*b*) at $\dot{M} \approx 0.3$–1.0 $M_\odot$ year^{-1}, the star becomes GR unstable during the helium-burning stage at $M_{\rm GR} \approx 2$–3.5×10^5 $M_\odot$, and (*c*) in an extreme case with $\dot{M} \approx 10$ $M_\odot$ year^{-1}, the star collapses during the hydrogen-burning stage at $M_{\rm GR} \approx 8.0 \times 10^5$ $M_\odot$.[9] Note that in the third regime, there is a discrepancy between the final stellar masses found by different groups: The largest discrepancy is less than a factor of three at the highest accretion rate (see Woods et al. 2019, their figure 10). The higher values of $M_{\rm GR}$ are caused by the positive entropy gradient in the realistic stellar structure. At the end of the stellar evolution, 60–80% of the total stellar mass is enclosed with the GR instability regime (Hosokawa et al. 2013, their figure 7). Because the unstable region is as hot as $T \sim 10^7$ K, the mass accretion rate after the onset of collapse is as high as $\sim$5,000 $M_\odot$ year^{-1}, which corresponds to a hyper-Eddington value of $\dot{M}/\dot{M}_{\rm Edd} > 10^6$ $(M_\bullet/10^5\ M_\odot)^{-1}$.

5.4. Massive Black Holes Via Runaway Mergers in a Dense Star Cluster

The previous subsection outlined the scenario in which the cooling and contracting of primordial gas in an ACH produces a massive seed BH via monolithic collapse into an SMS. Although this is the conventional wisdom for metal-free gas, the issue of whether the cloud may fragment at very high densities is not entirely settled. Furthermore, as mentioned in the introduction (see **Figure 3**) and in Section 5.1, when the gas collapsing in such a halo has some modest level of pre-enrichment by metals and/or dust, fragmentation is expected to occur. This fragmentation represents another possible pathway to massive seed BHs, or at least to IMBHs, which has received less attention to date than it deserves.

In particular, following the arguments outlined in Section 5.1, this fragmentation may occur at extreme densities—as high as $n_{\rm H} \sim 10^{11\text{–}13}$ cm^{-3} for very metal-poor clouds with $Z \gtrsim 10^{-5}$ $Z_\odot$ (see **Figure 7**), whereas the fragments have masses as low as $\sim$0.5 $M_\odot$, at least initially (corresponding to the Jeans mass at this density, shown also in **Figure 7**). The natural interpretation is therefore that an ultradense cluster of low-mass stars may result (Omukai et al. 2008, Devecchi & Volonteri 2009). Because of the extremely high density ($\sim 10^{9\text{–}11}$ $M_\odot$ pc^{-3}, i.e., orders of magnitude higher than the densities at the resolution limit of local stellar clusters; see below), such star clusters can undergo efficient runaway core collapse, which may consume a nonnegligible fraction of the stars, leading to a central IMBH with a mass of up to $\sim 10^4$ $M_\odot$.

The precise way a runaway collapse unfolds could take different shapes. Direct collisions between stars can occur over a timescale that is shorter than the lifetime of even massive stars (Katz et al. 2015, Yajima & Khochfar 2016, Sakurai et al. 2017), resulting in the formation of a single massive star. Note that this is a runaway process, because once stars collide, their masses and radii increase, accelerating the rate of subsequent mergers. Indeed, the basic features of this runaway have been elucidated in several studies addressing possible IMBH formation in the local Universe (Gültekin et al. 2004, Portegies Zwart et al. 2004, Freitag et al. 2006, Stone et al. 2017).

[9]For even more extreme cases with $\dot{M} \gg 10$ $M_\odot$ year^{-1}, which could be achieved in the merger of massive galaxies (Mayer et al. 2010) rather than in an ACH, the central region could directly collapse into a BH without forming a hydrostatic equilibrium structure such as an SMS (Haemmerlé et al. 2019, Mayer & Bonoli 2019).

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

We note, however, that all these works addressed star clusters with much lower densities than the densities possible in the high-z Universe. The bottom line is that stellar mergers could build a single SMS, essentially reproducing the pathway of the previous section, where such an SMS grew via accretion.

In the high-z context, a hybrid scenario is also possible. In particular, the newly born dense star cluster may be still embedded in a dense gas cloud. The dynamics of the runaway collapse itself may be aided by sudden significant gas inflows into the cluster (Davies et al. 2011). In this case, some of the protostars may also be accreting at high rates and have bloated envelopes, significantly increasing their geometric cross-section; gas dynamical friction also facilitates more efficient mergers (Boekholt et al. 2018, Reinoso et al. 2018). Additionally, the star–star mergers themselves may provide enough heating to keep a central star bloated and on track to the SMS regime (Tagawa et al. 2019). In these cases, a significant fraction ($\gtrsim$10%) of the stars may end up merging into a single SMS, and the resulting IMBH could reach masses as high as 10^4–10^5 $M_\odot$.

Finally, a variation on this possibility is that the cluster is somewhat less dense, and the stars are massive and short lived and produce remnant stellar-mass BHs before they merge. There will then be a population of stellar-mass BHs embedded in a dense gas cloud. Mergers between these BHs, aided by gas dynamical friction, could also build a more massive $10^{3–4}$ $M_\odot$ IMBH (Tagawa et al. 2016). Ryu et al. (2016) have addressed a similar setup, except they assume an even lower space density of the initial BH distribution (spreading the BHs over $\sim$100 pc rather than over $<$10 pc). They simulated the orbital motion of the BHs, taking into account gas drag on the BHs, and found that most initial BH configurations lead to the formation of a single massive BH in the center of the protogalaxy, reaching a mass of $10^{3–5}$ $M_\odot$ through hyper-Eddington growth rather than via BH–BH mergers.

For reference, a typical density in the core of a globular cluster is $\sim 10^3$ $M_\odot$ pc^{-3}, although many are much denser, and the densest known star cluster in the Milky Way, the Arches cluster, has a central density of $\approx 2 \times 10^5$ $M_\odot$ pc^{-3}. These densities are four to five orders of magnitude lower than that of the fragmenting clusters hypothesized at high redshifts but are measured on larger scales (Nguyen et al. 2018). Nuclear star clusters (Walcher et al. 2005) and even an isolated local ultracompact dwarf galaxy (Strader et al. 2013) reach surface densities at their half-light radii, which are similar to the densest globular clusters. However, their core densities can be much higher. For example, M32 has a central density, resolved by HST at $\sim$0.2 pc, in excess of $>10^7$ $M_\odot$ pc^{-3} (Lauer et al. 1998). M32 has a near-solar metallicity, and it also has a central SMBH with an estimated mass of 3×10^6 $M_\odot$ that could have generated the central Bahcall-Wolfe-like cusp (Merritt & Szell 2006). More generally, the above suggests that it may be possible to look for relics of the high-z star clusters in the local Universe. Such relics, assuming that they preserved their identities, would consist of the IMBH and a dense inner core of low-mass, extremely metal-poor stars, reaching $\sim 10^7$ $M_\odot$ pc^{-3} at $\sim 10^3$ AU (where the cluster did not have time to disperse in a Hubble time; Tagawa et al. 2019). Extrapolation of the stellar density in the nucleus of the Milky Way down to this scale is consistent with this value (Genzel et al. 2010). In summary, runaway mergers between stars and/or their remnant BHs, facilitated by the extreme densities and the presence of gas enveloping the young star cluster, represent a viable pathway to forming IMBHs, possibly via a stage of an SMS, but with masses expected to be somewhat below the most extreme massive seed BH produced in the other pathways.

5.5. Subsequent Growth and Cosmological Evolution

We have reviewed several possible formation pathways for seed BHs with masses of $10^2 \lesssim M_\bullet/M_\odot \lesssim 10^6$ in ACHs (or sub-ACHs). In this section, we briefly discuss the subsequent growth

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

of these seeds in the cosmological context of hierarchical structure formation. Two approaches have been utilized to model the population of growing BHs. The first is semianalytical modeling, in which BH seed formation, gas accretion, and mergers of BHs are modeled in a simplified way. This is an effective method for examining the statistical properties of BH populations and making theoretical predictions that can be directly compared with observations such as the quasar LF (see Section 2). The second approach is to use cosmological simulations of early galaxy formation. These simulations do not have adequate dynamical range for accurate global statistical predictions while resolving minihalos, but they can capture much more detailed properties of large-scale structure formation, galaxy mergers, the growth process of individual seeds into SMBHs, feedback associated with BH accretion, stellar radiation, and SN explosions.

Previous work employing semianalytical methods have elucidated the ingredients (e.g., seeding mechanisms/conditions and BH accretion physics) required to model the early BH population (e.g., Volonteri et al. 2003, Tanaka & Haiman 2009, Agarwal et al. 2013, Valiante et al. 2016, Ricarte & Natarajan 2018 and references therein). Based on two seed formation scenarios (Pop III remnants versus massive seeds with $M_\bullet \approx 10^5$ $M_\odot$), Tanaka & Haiman (2009) explored possible channels to explain the extremely rare high-redshift SMBH population. They varied key physical parameters such as the BH seeding fraction and the accretion duty cycle and found that models with the optimistic assumptions required to explain the SDSS SMBHs overproduce the mass density in lower-mass BHs by a factor of 10^2–10^3. They found that this overproduction can be avoided if seed formation stops or BHs accrete at lower duty cycles at $z \lesssim 20$–30 (such suppression is expected owing to the negative global feedback from the X-rays emitted by these BHs themselves; Tanaka et al. 2012). More recently, Valiante et al. (2016) developed semianalytical models that include additional physics related to BH seeding mechanisms including LW irradiation, chemical enrichment of halos, and cosmic reionization. They claim that a contribution of massive seeds with $M_\bullet \approx 10^5$ $M_\odot$ is needed to explain the most extreme SMBH population unless the typical mass of Pop III remnants extends to $\sim 10^3$ $M_\odot$. Interestingly, this implies that Pop III remnant BHs could play an important role in the formation of extremely massive BHs as well as the less-massive BH population.

In recent decades, large-scale cosmological simulations have been exploring the evolution of SMBHs and the coevolution of their host galaxies. Various feedback processes including SNe and AGN activity have been examined. Due to numerical limitations, cosmological simulations with a spatial resolution of $\sim O$ (kpc) must treat feedback effects with subgrid models instead of directly resolving physical processes in the nuclear region. There are many different simulation studies adopting different subgrid feedback models. These simulations have generally found that most (massive) seed BHs formed in protogalaxies hardly grow in mass via gas accretion because dense, cold gas is expelled by energetic SN feedback associated with star formation. Radiative/mechanical feedback caused by BH accretion is expected to produce a more modest effect (Dubois et al. 2013, 2015; Prieto & Escala 2016; Latif et al. 2018; Smidt et al. 2018; Smith et al. 2018). For example, Latif et al. (2018) focused on the growth of a 10^5 $M_\odot$ BH at the center of an ACH, marginally resolving the Bondi radius for cold gas with $T \approx 8{,}000$ K, and found that the accretion is completely shut off by SN feedback all the way down to $z \sim 6$ (see **Figure 12*a***). Habouzit et al. (2017) investigated the growth of seed BHs using large-scale cosmological simulations and found that most seeds cannot grow because of SN feedback until lower redshifts when those seeds settle in the centers of more massive galaxies. As a result of this SN feedback, they find that most seed BHs do not grow beyond $M_\bullet \approx 3 \times 10^3$ $M_\odot$ until they are in $M_* \approx 10^9$ $M_\odot$ galaxies in their strong feedback model (which is realistic in terms of reproducing galaxy observations). Their most massive BH reaches $M_\bullet \approx 3 \times 10^4$ (10^7) $M_\odot$ for their strong (weak) SN feedback model by $z \approx 6$.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

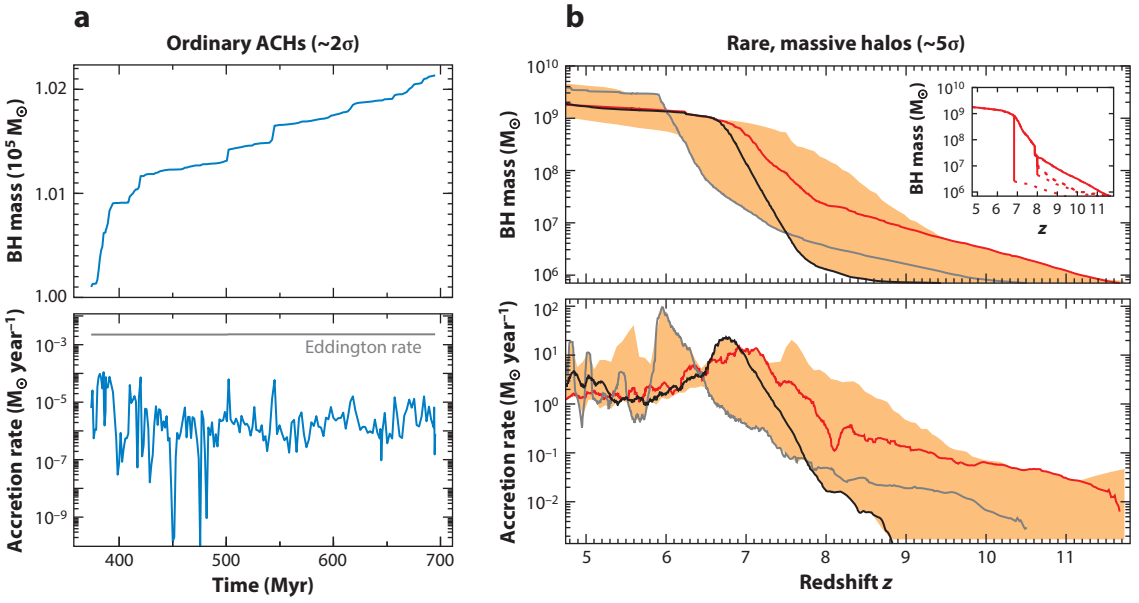

Figure 12

BH mass and the accretion rate versus redshift in two different simulations. (*a*) The growth of a 10^5 $M_\odot$ seed BH in a typical ACH halo, between $z = 12$ and $z = 7.5$. The host galaxy reaches 3×10^{10} $M_\odot$ at $z = 7.5$. (*b*) The evolution of seed BHs with the same mass but placed in the progenitors of more massive galaxies that reach $\sim 10^{12}$ $M_\odot$ by $z \approx 6$. The three lines represent three examples, and the orange band the full range in their simulated sample. In panel *a*, the seed BH never grows because of energetic SN and AGN feedback. In panel *b*, the seed BH is fed with cold gas streams supplied from large-scale structure at high accretion rates, and thus the BHs can grow to $M_\bullet \approx 10^9$ $M_\odot$ by $z \sim 7$. We caution that the numerical resolution and subgrid model treatments for feedback also differ in these simulations and could impact the overall BH growth efficiency. Panel *a* adapted with permission from figure 5 of Latif et al. (2018), and panel *b* adapted with permission from figure 3 of Di Matteo et al. (2012). Abbreviations: ACH, atomic cooling halo; AGN, active galactic nucleus; BH, black hole; SN, supernova.

It is very important to note that most simulations in which SN feedback quenches BH growth have focused on typical ACHs in $\approx 2\sigma$ regions of the Universe that will grow to $\lesssim 10^{11}$ $M_\odot$ by $z \approx 6$. However, the first SMBHs are likely hosted by rare $10^{12–13}$ $M_\odot$ halos (corresponding to 4–5σ fluctuations on these scales). In such massive halos, a sufficient gas supply would be maintained by cold gas streams from large-scale structure at $\dot{M} \sim 300$ $M_\odot$ year$^{-1}(V_{\rm circ}/200$ km s$^{-1})^3$ if all feedback processes are ignored. Cold gas streams feeding the center of the halo can exist only when thermal energy generated by shocks associated with virialization is quickly radiated away, otherwise a hot and diffuse medium fills the halo (Rees & Ostriker 1977, Silk 1977, Birnboim & Dekel 2003, Kereš et al. 2005, Dekel & Birnboim 2006). Large-scale cosmological simulations demonstrate that cold flows are not very susceptible to feedback from growing BHs and a high accretion rate is maintained until the mass of the galaxy reaches $\approx 10^{12}$ $M_\odot$, when the cold mode of accretion turns into the usual hot virialization mode and the gas supply to the nuclear region is strongly quenched (Di Matteo et al. 2012, Khandai et al. 2012). As a result, a seed BH with $M_\bullet = 10^5$ $M_\odot$ can grow up to $\gtrsim 10^9$ $M_\odot$ by $z \approx 6$ (see **Figure 12*b***), assuming a simple BH-feeding prescription of $\dot{M}_\bullet = \min(\dot{M}_{\rm Edd}, \dot{M}_{\rm B})$. This overall picture is consistent with several different mechanisms for seeding BHs of $\sim 10^{5–6}$ $M_\odot$ including intense LW radiation (Section 5.2.1), formation of dense shocked gas by colliding cold accretion flows (Section 5.2.2), strong streaming velocities (Section 5.2.3), and dynamical heating due to rapid halo mergers (Section 5.2.4). Similar conclusions have been reached in other studies, including by Li et al. (2007), Sijacki et al. (2009), and

Lupi et al. (2019), who find that SNe initially slow the growth, but once the galaxy is sufficiently massive, the BH can grow and reach 10^9 $M_\odot$ (but see Dubois et al. 2013, where depending on the AGN feedback model, the BH growth is severely self-regulated even in very massive halos). Therefore, further cosmological simulations in which the early growth of seed BHs takes place in peculiar environments self-consistent with their seeding models and the rarity of high-z SMBHs, as well as various feedback effects caused by SNe and AGN activities, are required.

There are several important limitations of current numerical simulations. First, most large-scale cosmological simulations do resolve the dynamics of DM/gas/stars on galactic scales at $\sim O$ (kpc), but the BH sphere of influence R_B is not resolved. Although some studies mentioned above marginally resolve the BH gravitational sphere of influence for neutral warm gas with $T \approx$ 8,000 K, the accuracy of their prescriptions for energy and/or momentum feedback injected in unresolved regions still remains uncertain. Additionally, the threshold density above which gas turns into stars is typically set to $n_H \approx 1$–100 cm^{-3}. This is orders of magnitude lower than the density in star-forming regions in the local Universe. The efficiency of stellar and SN feedback must strongly depend on this star-formation density parameter. For example, if the threshold is set to a very low value, dense and cold gas clouds are disrupted because gas consumption and feedback strength are artificially overestimated. This could lead to the very inefficient growth of BHs seen in many cosmological simulation studies and could suppress possible short-duration super- or hyper-Eddington accretion phases (we note that most previous large-scale cosmological simulations do not include the possibility of super-Eddington accretion, which should be addressed in future works). Another important piece of subgrid physics is how the orbits of massive BHs are handled during the mergers of their host galaxies. A common approach is to simply assume that once two BHs are sufficiently close, they merge instantly, and the merger remnant is moved to the center of mass of the host galaxy. In reality, the unresolved inspiral of massive BHs can be inefficient and take a significant fraction of the Hubble time; during this inspiral their growth is suppressed by the lower ambient density and high orbital speed (see, e.g., Gabor et al. 2016 and references therein). Cosmological simulations that self-consistently connect with much higher-resolution noncosmological simulations and are able to resolve the orbits of massive BHs to within the GW–emitting regime will likely be needed in the future for more accurate predictions.

In summary, most massive seed BHs formed in early protogalaxies at $z \gtrsim 10$ would hardly grow via gas accretion because energetic SN feedback quickly evacuates gas from the nuclear region. However, a small minority of seeds that were born in highly biased regions of the Universe could be fed with intense cold accretion streams through large-scale cosmic filaments. Additional theoretical work on the link between the environments required for BH seeding and the assembly history of SMBH host galaxies is needed.

6. ALTERNATIVE BLACK HOLE FORMATION CHANNELS

In the absence of a definitive conclusion that early BHs formed via one or more of the astrophysical scenarios above, it is prudent to keep an open mind to alternative possibilities. Also, the *James Webb Space Telescope* (JWST) and other instruments will have the observational capability of detecting more distant and more massive SMBHs, which can only be produced in exotic scenarios. Here, we briefly review ideas invoked to form (perhaps some of the) early massive BHs.

6.1. Primordial Black Holes

The notion that PBHs may have formed in the early Universe was suggested over 50 years ago (Zel'dovich & Novikov 1967, Hawking 1971). Many specific PBH formation mechanisms have been developed, based on large density fluctuations that can decouple from the cosmic expansion

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

and collapse into BHs. Such overdensities could be produced in many different ways, including in phase transitions, a temporary softening in the equation of state (reducing pressure), quantum fluctuations, or specific designer models of inflation leading to a narrow peak in the fluctuation power spectrum (see, e.g., a review by Carr 2006). Equating the cosmological background density at cosmic time t with the density of a Schwarzschild BH yields a characteristic PBH mass tracking the increasing horizon mass,

$$M_{\rm PBH} \approx \frac{c^3 t}{G} \approx 10^5 \left(\frac{t}{1\rm s}\right) \rm M_\odot. \quad 9.$$

Interest in PBHs has also come from many contexts, such as the various consequences of the Hawking radiation from small PBHs ($\lesssim 10^{15}$ g) that have evaporated and are currently evaporating (resulting in, e.g., a γ-ray background; Page & Hawking 1976), the possibility that PBHs make up most of or even all the DM (e.g., Carr et al. 2016), or the idea that they may account for recent Laser Interferometer Gravitation-Wave Observatory (LIGO) detections of mergers between stellar-mass BHs in the local Universe (Bird et al. 2016).

Most interesting for this review is the possibility that PBHs as massive as 10^5 $\rm M_\odot$ exist and take on the role of massive seeds at high redshifts (e.g., Bean & Magueijo 2002, Dolgov et al. 2009, Dolgov 2018). Such massive PBHs form relatively late (i.e., at $t \sim 1$ s; Equation 9), near the epoch of electroweak decoupling (although earlier production during inflation is possible in some models; Khlopov 2010, Belotsky et al. 2019), and therefore could potentially disturb the successful predictions of Big Bang nucleosynthesis, introducing small (perhaps as large as ∼1%) inhomogeneities in the He abundance (Carr & Silk 2018). The strongest constraints on such massive PBHs, however, come from the angular power spectrum of CMB temperature and polarization anisotropies. Massive PBHs begin to accrete gas efficiently after cosmic recombination (Miller 2000). The corresponding radiation preionizes and heats the IGM (Ricotti et al. 2008). The accompanying impact on the CMB anisotropies is subject to uncertainties about the nature of the accretion (e.g., typical angular momentum and corresponding mode of accretion) and the emerging radiation (overall radiative efficiency and spectrum). The effect is strongest when the radiation is assumed to arise from a disk and yields a limit of

$$\rho_{\rm PBH} \lesssim 1.1 \times 10^3 \left(\frac{10^5\ \rm M_\odot}{M_{\rm PBH}}\right)^{1.6} \left(\frac{0.01}{\lambda}\right)^{1.6} \rm M_\odot\, Mpc^{-3} \quad 10.$$

on the comoving mass density of these PBHs. This limit was derived by Poulin et al. (2017) based on radiatively inefficient advection-dominated accretion flow (ADAF) models and includes an uncertain fudge factor λ (≈0.01) by which the disk accretion rate is reduced (by winds and outflows) compared to the Bondi–Hoyle–Lyttleton rate. A model adopting radiatively even less efficient spherical accretion (Ali-Haïmoud & Kamionkowski 2017) finds a ∼100 times weaker limit. Equation 10 represents a tiny fraction ($\approx 3 \times 10^{-8}$) of the comoving DM density and only 0.3% of the total present-day SMBH mass density (e.g., Yu & Tremaine 2002). Nevertheless, it corresponds to only a weak upper limit of $\rho_{\rm PBH}/M_{\rm PBH} \approx 0.01$ $\rm Mpc^{-3}$ on the comoving number density of such BHs, which is comparable with the present-day galaxy number density and is $\sim 10^{6–7}$ times larger than the abundance of bright quasars at $z \approx 6$–7. Therefore, at least by their abundance, massive PBHs remain viable as seeds of rare early quasars. Equation 10 also shows that the abundance of smaller PBHs allowed by CMB constraints increases rapidly (because these smaller BHs accrete less efficiently from the IGM). In the $1\ \rm M_\odot \lesssim M_{\rm PBH} \lesssim 100\ M_\odot$ range (where CMB constraints are weak), a limit is provided by the (lack of) weak gravitational lensing of type Ia SNe (Zumalacárregui & Seljak 2018). This limit is still weak and allows up to ∼30% of the DM to be composed of such stellar-mass PBHs. The strongest current limit comes from the number

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

of LIGO events, which suggests that 1 $M_\odot \lesssim M_{\rm PBH} \lesssim 300\ M_\odot$ BHs cannot make up more than ∼1% of the DM (Ali-Haïmoud et al. 2017).

6.2. Dark Matter–Powered Stars

Another possibility is that some first-generation stars in the Universe were so-called dark stars (Spolyar et al. 2008). Dark stars are similar to normal metal-poor stars except that they are powered by DM annihilation (see a recent review by Freese et al. 2016). The energy release from DM self-annihilation can replace nuclear fusion as the dominant energy source and maintain a hydrostatic structure as long as the DM fuel lasts. Annihilation converts most of the rest-mass of the DM particles to energy that can heat the star, compared to the 0.7% efficiency of baryonic fusion. As a result, DM typically makes up a small fraction ($\lesssim 10^{-3}$) of the total mass of the dark star.

A dark star can exist only given the following conditions: (*a*) The DM density within the star is sufficiently high for annihilation to dominate over other heating processes (nuclear fusion, gravitational contraction) and (*b*) the annihilation products must deposit their energy in the stellar envelope, where it must thermalize rather than escape the star. One of the most popular (and best motivated) DM candidates is a weakly interacting massive particle (WIMP). In many models the WIMP is its own antiparticle and self-annihilates with a cross section $\langle \sigma v \rangle \approx 10^{-26}\ {\rm cm^3\ s^{-1}}$ determined by the strength of weak interactions, which yields the observed DM density (Jungman et al. 1996). A natural place in which dark stars may be expected to form is in regions of high DM density in the early Universe, at the cores of DM minihalos. Assuming a Navarro–Frenk–White (NFW) profile in these early halos, consisting of WIMPs, satisfies the first condition above. Whether the second condition is satisfied depends on the specific types of annihilation products (e.g., photons, electrons/positrons, or neutrinos) and their energy spectrum. For typical WIMP models, model stellar atmospheres have been found to be opaque to these annihilation products (or their secondary products as they cascade through a sequence of photon-electron-positron conversions; see Freese et al. 2016). More generally, the requirements for dark star existence have been met generically for a wide range of WIMP properties and DM halo density profiles (Freese et al. 2009).

The main difference between normal metal-poor massive stars and dark stars is the latter are much puffier (several astronomical units in radius), and therefore owing to their large surface area and correspondingly low surface temperature ($\lesssim 10^4$ K), they emit primarily IR radiation. Similar to the case of the rapidly accreting protostars discussed in Section 5.3.2 above, this avoids UV feedback shutting down their accretion and limiting their masses: Dark stars can continue to grow. However, for dark stars, the external gas accretion rate need not be very high to maintain their large size. For example, at the accretion rate of $10^{-3}\ M_\odot\ {\rm year^{-1}}$, DM annihilation can keep the star bloated and its effective temperature near $\sim 10^4$ K, so that in 10^8 years, a $\sim 10^5\ M_\odot$ SMS is built (e.g., Rindler-Daller et al. 2015). In principle, dark stars can continue to grow to even higher masses up to $10^7\ M_\odot$, given a sufficiently high $\dot{M}$, as long as the original DM fuel lasts and/or if the DM within the star is replenished by the capture of new DM particles. In this case, they may reach luminosities high enough to be detectable by the Near-Infrared Camera (NIRCam) on JWST in megasecond exposures (e.g., Ilie et al. 2012). Ultimately, once their DM fuel is exhausted, the most massive dark stars will collapse to BHs via a general relativistic instability without ever going through a phase of nuclear fusion.

6.3. Heating by a (Primordial) Magnetic Field

As argued elsewhere in this review, the key ingredient of massive BH seed formation is for the collapsing gas to avoid fragmentation, which in turn requires avoiding efficient H_2 cooling and remaining at temperatures near $T \sim 10^4$ K. This condition can be satisfied by either reducing

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

cooling or, alternatively, invoking an extra source of heating, beyond the compressional and shock heating that is typically included in studies of gas collapse in protogalaxies. In principle, a sufficiently strong [$\sim$1 nano-Gauss (nG)] primordial magnetic field (PMF) could help provide the required heating (Schleicher et al. 2009, Sethi et al. 2010).

Several mechanisms have been proposed to produce a global PMF, with a comoving field strength of order 1 nG (quoted here and throughout this subsection on the scale of 1 Mpc), during inflation and/or during various phase transitions in the early Universe [see, e.g., extended reviews by Widrow (2002), Subramanian (2016) and references therein]. If present, the weak seed PMF can be amplified by flux-freezing inside a collapsing primordial gas (Maki & Susa 2007, Sethi et al. 2008, Schleicher et al. 2009). Dynamo effects should also help amplify small initial seed fields (Schleicher et al. 2010a); simulations resolving turbulence have confirmed this and suggested that magnetic fields can become dynamically important (e.g., Turk et al. 2012).

Several studies have pointed out that a seed PMF can therefore affect the fragmentation properties of the first protogalaxies (e.g., Machida & Doi 2013, Latif et al. 2014b). First, $\sim$1-nG seed magnetic fields can elevate the Jeans mass and delay collapse until past the atomic-cooling threshold (Schleicher et al. 2009, Sethi et al. 2010). Second, heating by ambipolar diffusion or decaying turbulence can dominate H_2 cooling and keep the gas warm and the gas accretion rate high. In one-zone models, Sethi et al. (2010) find that ambipolar diffusion dominates, and the critical PMF value that keeps the gas at $T \sim 10^4$ K is $\sim$3 nG.

This critical magnetic field strength is somewhat higher than recent upper limits from the CMB. PMFs leave several distinct signatures in the CMB anisotropies. From the angular power spectra, the *Planck Surveyor Satellite* finds a limit of 2 nG on a nearly scale-invariant PMF (Planck Collab. 2016). This limit is strengthened by a factor of $\sim$2 by including the impact of the PMF on the ionization history (Kunze & Komatsu 2015, Paoletti et al. 2019) or by combining *Planck* and the South Pole Telescope data to use the small-scale B-mode polarization power spectrum, which is sourced by the PMF and survives damping well past the Silk scale (Zucca et al. 2017). Ultrafaint dwarfs, whose formation at high redshift would be suppressed for large PMFs, give an even stronger limit of $\sim$0.5 nG (Safarzadeh & Loeb 2019).

Overall, this suggests that for a PMF near its maximum allowed amplitude of $\sim$1 nG, H_2 cooling can be fully suppressed in rare $\gtrsim 3\sigma$ regions of the spatially fluctuating B–field. However, it is worth noting that even weaker fields can be important, because they can reduce the LW flux ($J_{\rm crit}$) required to disable H_2 cooling; Van Borm & Spaans (2013) find a factor of 10 reduction for $B = 2$ nG.

6.4. Massive Black Holes from Collisional Dark Matter

A handful of well-documented issues with galaxy formation in the ΛCDM model (e.g., the so-called cusp/core, missing satellite, and too-big-to-fail problems; reviewed recently by Bullock & Boylan-Kolchin 2017) have motivated several proposals to change the particle properties of CDM. In general, these modifications are designed to reduce the fluctuations on small scales (below $\sim$1 Mpc, or $\sim 10^{11}$ $M_\odot$). Such reductions can be dramatically important for early SMBH formation. In the CDM paradigm, the smallest objects collapse first, and they subsequently merge together to form larger objects. It then follows that the loss of small-scale power modifies structure formation most severely at the highest redshifts; in particular, the number of self-gravitating objects at high redshift is reduced. For example, in the context of warm dark matter (WDM) models, Barkana et al. (2001) have shown that a WDM particle mass significantly below 1 keV would make it difficult to form any DM halos at $z > 7$ in order to host a high-redshift quasar. Likewise, the highest-redshift $z = 10$ galaxies detected in the Hubble deep fields behind strong

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

lensing clusters would be hard to explain in WDM models with this particle mass (Pacucci et al. 2013). By contrast, a particle mass near 1 keV would erase the population of high-z minihalos and delay the onset of structure formation to commence with ACHs (Dayal et al. 2017). This would help create several metal-free ACHs that are conducive to massive BH formation.

In the case of SIDM (Spergel & Steinhardt 2000), two other interesting effects arise. First, due to the self-interactions (compared to much weaker purely gravitational interactions), the gravothermal catastrophe sets in on a much shorter timescale and can result in the relativistic collapse of the core of a DM halo into a BH (see, e.g., Koda & Shapiro 2011 for a detailed discussion). Balberg & Shapiro (2002) found that $\sim 10^6$ $M_\odot$ BHs in the cores of massive ($\sim 10^{12}$ $M_\odot$) halos can form by $z \sim 10$ as a result, although they used a cross section $\sigma = 5$ cm^{2}g^{-1} that is now ruled out. The interaction cross section is limited to $\sigma < 0.6$ cm^{2}g^{-1}; this comes from the displacement of the gas with respect to the centroid of the total mass in the merging subcluster component of the Bullet Cluster, which shows that unlike the baryons, the DM component has not been slowed down by collisions (Randall et al. 2008). A twist on the SIDM idea, however, is that the bulk of DM is normal CDM, but a small fraction is SIDM with a large cross section. This hybrid model avoids essentially any constraint from the overall behavior of DM, and forming $\sim 10^6$ $M_\odot$ BHs by $z \sim 10$ appears feasible in such strong-SIDM subcomponent models (Pollack et al. 2015, Choquette et al. 2019).

A second feature of SIDM is that a preexisting BH can accrete efficiently. This is because SIDM is presumed not to radiate, so the Eddington limit does not apply, and because scattering allows a rapid diffusive refilling of the loss-cone; thus efficient accretion continues even when the SIDM mean free path is larger than the Bondi radius (Ostriker 2000). This could help in growing massive BHs, starting from stellar-mass seed BHs formed in the usual way (i.e., a remnant of a massive star), located at the dense core of their DM halos.

7. FUTURE OBSERVATIONAL DIAGNOSTICS

Existing $z \approx 6$–7 quasar observations only probe the most massive SMBHs, representing the tip of the iceberg of the high-redshift BH population. In order to better understand early BH formation and growth, it will be important to characterize a much wider mass range and to go to higher redshifts. Fortunately, a variety of planned or proposed instruments, such as JWST, the *Laser Interferometer Space Antenna* (LISA), and the *Lynx X-ray Observatory* (*Lynx*) will make this possible.[10] In this section, we discuss the most promising observational probes, including both direct detections and indirect methods at high redshift, as well as fossil evidence from BHs in the local Universe.

7.1. Direct Observations

Current observations of SMBHs at $z \approx 6$–7 are unlikely to contain information on SMBH seeding. This is because, even for massive seeds, SMBHs have grown many orders of magnitude in mass, erasing the memory of seed formation and early accretion. Furthermore, whatever combination of seeding and accretion produced these rare SMBHs is likely to be a small unrepresentative fraction of all the BHs born at much higher redshifts. To distinguish seed models, and to diagnose the IMF of early BHs at their birth and their subsequent growth, it will be important to observe BHs near (or below) the masses predicted for massive seeds, $\sim 10^5$ $M_\odot$, at redshifts $z \gtrsim 10$, where they are expected to form. For reference, we note that the flux from a 10^5 $M_\odot$ BH at $z = 10$, assuming

[10] See **www.jwst.nasa.gov**, **www.lisamission.org**, and **www.lynxobservatory.com**.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

that it shines at the Eddington limit and has the same spectral shape as typical lower-z quasars, would be $\sim$0.5 nJy in the near-IR (observed at 1 μm) and $\sim 3 \times 10^{-19}$ erg s^{-1} cm^{-2} in the soft X-ray (observed at 1 keV) bands. Directly detecting these BHs will be challenging, but doing so should be possible in megasecond exposures with JWST's NIRCam and in the soft X-ray bands with a new, sensitive instrument such as the planned *Lynx* telescope, which can reach a sensitivity of $\sim 10^{-19}$ erg s^{-1} cm^{-2} in 4 Msec. Searching areas of strong gravitational lensing behind massive clusters could improve these detection limits by an order of magnitude, although only in small solid angles. Once these high-z point sources are detected, the next question is whether there is any clear signature of their origin. Three distinctive features of massive seed BHs born from pristine gas in ACHs are (*a*) obscuration by an unusually large gas column, placing the bulk of their emergent flux in the IR and X-ray bands; (*b*) the lack of any (or very little) metals in their spectra; and (*c*) the lack of an appreciable host galaxy.

With the sensitivities above, the IR and X-ray LFs of accreting high-redshift BHs at $z = 10$ should be possible to measure over 2–3 orders of magnitude in flux, which will help characterize the BH population and constrain seeding and evolution models. Stellar-mass seeds are generally thought to be much more abundant than massive seeds: Cosmological hydrodynamical simulations and semianalytic calculations find a cumulative Pop III stellar density of $\approx 3 \times 10^5$ M$_\odot$ Mpc^{-3} forming by $z \approx 6$ (see, e.g., Wise et al. 2012, Visbal et al. 2015). If the Pop III IMF is top-heavy as predicted by Hirano et al. (2014), a significant fraction of this total Pop III mass density will end up in BHs. Whereas the abundance of massive seeds is highly uncertain, they are expected to be much rarer. For massive seeds created through strong LW feedback from a neighboring galaxy, an extremely optimistic value of the critical LW flux ($J_{\rm crit} \approx 100\ J_{21}$) leads to a number density of $\sim 10^{-5}$ Mpc^{-3} (Dijkstra et al. 2014). Including radiation from satellite galaxies in the same halo, and/or including other effects mentioned in Section 5.2, the predictions rise as high as $\sim 10^{-3}$ Mpc^{-3}. This value is reached under optimistic assumptions about (the lack of) metal enrichment in the synchronized halo pairs scenario (Visbal et al. 2014b). Wise et al. (2019) find a similar value due to dynamical heating of rare ACHs, although in their simulations this value only applies in highly biased regions (overdense on large scales) and the global average abundance is $\sim(10^{-6}$–$10^{-7})$ Mpc^{-3}. Reaching even higher values requires unusual or ad hoc assumptions. For example, if $J_{\rm crit}$ were 1–2 orders of magnitude lower than expected, then the abundance could reach up to $\sim 10^{-1}$ Mpc^{-3}, corresponding to a large fraction ($\sim$10%) of all ACHs at $z \approx 10$ (see Habouzit et al. 2016 and references therein). These values, however, are still orders of magnitude below the abundance of Pop III seeds. For reference, we note that the space density of 10^{-3} Mpc^{-3} at $z \approx 10$ would correspond to $\approx$1 arcmin^{-2} per unit redshift, which is sufficient to collect a statistical sample with IR/X-ray surveys.

Recent semianalytic models (Ricarte & Natarajan 2018) made predictions for the high-redshift LFs for stellar and massive seeds. At sufficiently high redshifts ($z \gtrsim 10$), the LFs diverge and depend strongly on the seed model implemented. In particular, the bright end is strongly impacted by the seed mass: stellar BHs whose growth is capped by both the approximate Eddington rate and radiative feedback (assumed to maintain the $M_\bullet$–σ relation) cannot produce as massive and luminous SMBHs by $z = 10$ as can the brightest massive seeds. However, the faint end is strongly affected by the abundance of seeds (with stellar seeds producing a much higher number of quasars at, e.g., *Lynx*'s detection threshold). However, it must be kept in mind that the assumed accretion properties of high-redshift BHs strongly impact the LF. Tanaka & Haiman (2009) find that the BH mass function depends on a combination of accretion, duty cycle, and seeding fraction. These degeneracies will likely make it difficult to characterize the population of high-redshift BHs from the LFs alone.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

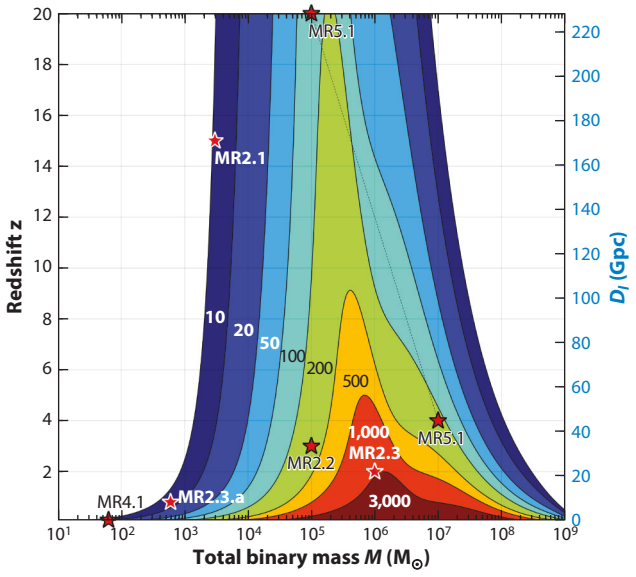

Figure 13

Contours of constant signal-to-noise for detection of gravitational waves from massive BH mergers by LISA, as a function of redshift and binary mass (assuming a mass ratio $q = 0.2$). Red stars indicate reference binaries used to define MRs, which include the detection of 10^3 $M_\odot$ binaries to $z = 15$ (MR2.1) and the accurate characterization of the postmerger waveform from 10^5 $M_\odot$ binaries to $z = 20$ (MR5.1). Abbreviations: BH, black hole; LISA, *Laser Interferometer Space Antenna*; MR, mission requirement. Figure adapted from Amaro-Seoane et al. (2017).

7.1.1. Gravitational waves and other transients. The IR and X-ray LFs of BHs will inform us mainly about gas accretion onto high-redshift BHs. This will be strongly complemented by future GW measurements with the space-based detector LISA, planned for launch in 2034. LISA will have the sensitivity to detect BH mergers for masses from $\sim 10^{4–7}$ $M_\odot$ out to redshifts beyond $z \approx 20$ (see **Figure 13**). Simultaneously probing BH mergers and accretion will give us a complete picture of BH growth in the early Universe.

Theoretical predictions of LISA event rates for merging massive BHs have been derived from numerous semianalytical models as well as more recently from cosmological galaxy evolution simulations. The predictions span a wide range, from no detectable events in the most pessimistic case up to several tens of events per year per unit redshift at $z = 10$, and a few events per year even at $z \gtrsim 15$ in the most optimistic models (for recent predictions see, e.g., Klein et al. 2016, Hartwig et al. 2018a, Ricarte & Natarajan 2018; for earlier studies that included the $z \gtrsim 10$ Universe see, e.g., Sesana et al. 2007, Tanaka & Haiman 2009 and references therein). The most pessimistic scenario is for stellar-mass BHs to stay below LISA's detection threshold and for massive seed BHs to form inefficiently or to be unable to promptly merge when their hosts merge. The largest rates arise when abundant stellar-mass BH seeds are assumed to grow above LISA's $\gtrsim 10^{3–4}$ $M_\odot$ detection threshold by $z \sim 10$ or in models in which massive seeds are assumed to form efficiently in a large fraction of ACHs. Aside from the total event rates, the mass spectrum is expected to depend on the seeding and should help disentangle models. LISA data can indeed distinguish between a wide variety of models that include different prescriptions for seed formation, feedback, accretion efficiency, and accretion geometry (Sesana et al. 2011). LISA may also be able to probe SMS star formation itself by measuring mergers of the remnants of binary SMSs (Hartwig et al. 2018a) or

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

directly detecting their formation process, via either the burst accompanying the collapse of a rotating SMS into a single massive BH (Saijo & Hawke 2009, Shibata et al. 2016) or the breakup of such an SMS producing two massive BHs, which subsequently inspiral and merge (Reisswig et al. 2013).

Although LISA will probe a BH mass range important for probing massive seeds, significantly smaller stellar seed mergers cannot be detected. The proposed interferometry mission *DECi-hertz Interferometer Gravitational wave Observatory* (DECIGO; Sato et al. 2017) as well as ongoing and future experiments based on quantum interferometry with cold atoms, such as the terrestrial MAGIS-100 (100-meter-long Matter-wave Atomic Gradiometer Interferometric Sensor; Coleman 2018) or the AEDGE (*Atomic Experiment for Dark Matter and Gravity Exploration*) satellite (El-Neaj et al. 2020), would fill in the gap between LIGO and LISA. This is warranted especially because the LIGO discoveries have already uncovered stellar-mass BHs more massive than had been expected (currently up to 85 $M_\odot$; Abbott et al. 2019), raising the possibility that the mass function extends to even higher masses.

In addition to GWs, there are several other transient observables that will be important for constraining the high-redshift BH population. Surveys with JWST or WFIRST will have the ability to see the potentially extremely luminous PISNe from Pop III stars with masses of $M_\star \sim$ 140–260 $M_\odot$ to redshifts as high as $z \sim 30$ (Kasen et al. 2011). Less massive but potentially much more numerous PISNe from lower-mass ($M_\star \sim$ 90–140 $M_\odot$) Pop III stars could still be visible to $z \sim 10$ (Smidt et al. 2015). Observing these explosions would put constraints on the abundances and IMFs of Pop III stars (and thus the prevalence of light BH seeds). However, Hartwig et al. (2018b) pointed out that the identification of PISNe and differentiation from other sources that could have a similar photometric signature, such as AGNs or high-z galaxies, are very challenging; in fact, the optimal strategy would require at least 50,000 different fields of view with an exposure time of ≈600 s for each field to detect one PISN at $z < 7.5$. It may also be possible to directly detect the explosions of SMSs, yielding information on massive seeds. Nonrotating SMSs in a narrow mass range near 55,000 $M_\odot$ have been found to undergo explosive nuclear burning and leave no remnant (Chen et al. 2014). Whalen et al. (2013) simulated the light-curves of these rare ultrabright thermonuclear SNe and found that they should be detectable to *Euclid* and JWST to $z \sim 20$. However, most SMSs are expected instead to rotate and produce a nuclear BH with an accretion disk, which launches a relativistic jet or outflow, analogous to long gamma-ray bursts (GRBs). Numerical work suggests that the jets could break out from the bloated stellar envelope, producing an ultralong GRB lasting for 10^4–10^6 s, as well as a bright optical afterglow detectable with *Euclid*, WFIRST, and JWST out to $z \sim 20$ (e.g., Matsumoto et al. 2015, 2016). Tidal disruption events of stars formed in the dense accretion disk around a newly born massive seed BH could occur, and the bright X-ray and radio emission from the jets from these events could be detected out to $z \sim 20$ and diagnosed by their long duration of $\sim 10^{5-6}(1 + z)$ s (Kashiyama & Inayoshi 2016).

7.1.2. Spectral signatures. Another approach is to identify accretion onto massive seeds shortly after their formation from their spectra. Due to the monolithic collapse of hydrogen gas required to form massive seeds, a newly formed massive BH seed should still be buried in a very large column density of gas. The impact of the corresponding obscuration has been investigated in 1D radiation-hydrodynamic simulations (Pacucci et al. 2015a). These simulations find strong X-ray (0.1–100 keV) and submm/IR (1–100 μm) emission, with a large gap in the spectrum at wavelengths below the Lyα line owing to absorption and reprocessing to lower-wavelength radiation. Using similar simulations, Pacucci et al. (2016) and Natarajan et al. (2017) have predicted that JWST will have the sensitivity to observe newly formed massive seeds and proposed a set of

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

color–color cuts that can be used to distinguish these objects from high-redshift galaxies and low-redshift contaminants on the basis of their very red IR colors. Additionally, the combined spectra of growing BH seeds and their host galaxies were predicted by Valiante et al. (2018). They find that accreting high-redshift seeds will be detectable with JWST but that it will be difficult to distinguish stellar versus massive BH seeds. We also note that theoretical predictions for the spectra of SMSs (before BH formation) suggest that they could also be detected by JWST (Hosokawa et al. 2013, Surace et al. 2018).

An additional spectral signature of massive seeds was identified by Dijkstra et al. (2016b). The high column density of pristine gas and the absence of dust lead to numerous Lyα scatterings. This creates ideal conditions for pumping of the $2p$ level of atomic hydrogen, which results in 3-cm (rest frame) maser emission from the $2p \rightarrow 2s$ transition. This bright line could be observed by the Square Kilometer Array (SKA).

Observing the characteristic spectrum of a massive BH seed buried by a very high column of gas and/or producing 3-cm maser emission could provide a smoking gun for the presence of a massive seed BH. However, one major challenge is that these observations require the seed BH to have formed very recently ($z \lesssim 6$–10). The abundance of massive seeds therefore needs to be close to the highest predictions to have a chance of catching their formation within a narrow time window.

A related spectral diagnostic is the Lyα line emitted from an accreting massive seed BH. Dijkstra et al. (2016a) model the Lyα line profile in this scenario and find line offsets and widths exceeding $\sim$1,000 km s^{-1}. This is higher than typical Lyα-emitting galaxies and, thus, may be an interesting observational discriminant.

Another interesting spectral signature is the presence of strong HeII recombination lines (with the strongest line at a rest-frame wavelength of 1640 Å). This line is expected, and should be detectable by JWST to $z \approx 10$, from both Pop III stars and accreting massive BHs (Tumlinson & Shull 2000, Oh et al. 2001). An important difference is that the Pop III stellar lines are expected to be narrow, because of the absence of strong Wolf-Rayet-type winds, whereas lines produced in the vicinity of an accreting BH should generally be much broader. The claimed detection (Sobral et al. 2015) in the bright $z = 6.6$ Lyα emitter CR7 of a strong HeII 1640 Å line without corresponding metal lines was interpreted by several authors as evidence for either Pop III stars or a massive seed BH (although the narrow observed width, which favors a stellar origin, received very little attention). Although a subsequent analysis found any HeII line much weaker, and the metallicity higher than originally believed (Shibuya et al. 2018), CR7 served as an intriguing case study. Future detection of a similarly strong but broad HeII line without metals could signal an SMS that recently collapsed to a BH. A strong and narrow HeII line would likewise indicate a significant population of Pop III stars, which are predicted to form concurrently with the massive seed BH in models in which the seed is forming in a galaxy significantly above the atomic-cooling threshold (Inayoshi et al. 2018).

7.1.3. Black hole host galaxies. Low-z SMBHs follow tight correlations with their hosts, which are widely believed to result from feedback processes (Kormendy & Ho 2013). Therefore seed BHs will not generally be born on this relation but will rather settle onto the relation over time (typically "from above," with BH masses initially above the relation). The high-z evolution of the low-mass end of well-known BH–host relations should therefore contain information on the earliest seeds and their growth (e.g., Volonteri & Natarajan 2009, Pacucci et al. 2018).

More specifically, a discussed above, the standard formation scenario for massive seeds requires ACHs with no prior star formation and a suppressed H_2 abundance to avoid metal cooling and gas fragmentation. Thus, DM halos hosting massive seeds will initially have an obese-BH galaxy

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

stage (Agarwal et al. 2013), where there is either no appreciable host galaxy at all or BH mass strongly dominates over the stellar mass, and the accretion onto the BH outshines the stellar component of the host galaxy. An important question is how long a typical newborn massive BH will stay unusually obese before its host grows more massive. This was investigated by Visbal & Haiman (2018), who tracked several thousand ACHs in a cosmological *N*-body simulation. It was found that before they can grow by an order of magnitude in mass, essentially all massive seeds have BH mass–to–stellar mass ratios significantly higher than stellar-mass seeds that have grown to the same mass. A promising strategy to diagnose massive seeds will therefore be to measure their accretion rates with next-generation X-ray telescopes such as *Lynx* and compare their host galaxies' stellar properties using data from either JWST or 30-m-class ground-based telescopes. Strong X-ray sources that are not accompanied by any detectable stellar host component will be strong massive seed candidates. Even in a few rare cases where the ACH host of a massive seed promptly merges with a nearby massive galaxy, Visbal & Haiman (2018) found that the ACH remains offset from this galaxy at a distance (a few kiloparsecs) that can be resolved with JWST or *Lynx*. In short, the tell-tale evidence for a massive seed BH is the absence of a host galaxy or a galaxy that is offset by a few kiloparsecs. In a variant of this picture, the X-rays from the growing massive seed BH trigger the formation of a small cluster of Pop III stars, which could lead to characteristic spectral signatures (Barrow et al. 2018).

7.2. Indirect Observations

In addition to direct detection of individual objects, there are promising indirect observations that can shed light on the high-redshift BH population. One possible approach is to measure the heating and partial ionization of the IGM caused by X-rays produced during BH accretion. When an X-ray ionizes a hydrogen or helium atom, a large amount of energy is imparted to the escaping electron. This energetic electron then interacts with the surrounding gas leading to heating and secondary ionizations. For a nearly neutral IGM, a significant fraction of X-ray energy goes into both heating and ionization (for detailed calculations, see Furlanetto & Stoever 2010). As the ionization fraction increases, a larger percentage of X-ray energy is deposited as heat, making X-rays less efficient than UV photons at ionizing the IGM to an ionization fraction close to unity.

One probe of such early "preionization" is via CMB temperature and polarization anisotropies. The optical depth to electron scattering, τ_e, provides an integral constraint on the ionization of the IGM. Due to the increased density of the Universe at high redshift, even a relatively small prereionization can contribute significantly to τ_e. Thus, CMB observations can be utilized to probe the total accretion of seed BHs in the early Universe (Madau et al. 2004, Ricotti & Ostriker 2004, Ricotti et al. 2005). Because τ_e is only one number, it is not possible to disentangle the quantities impacting a BH-driven prereionization (e.g., spectral properties, abundance, duty cycle, radiative efficiency, etc.); τ_e measurements only yield upper limits on early preionization. With reasonable (but uncertain) assumptions, Visbal et al. (2015) show that if most Pop III remnant BHs accrete radiatively efficiently near the Eddington limit, the resulting prereionization would violate the electron optical depth constraints from *Planck*.

In principle, the shape of the large angular-scale CMB polarization power spectrum contains information on the evolution of the ionized fraction $x_e(z)$ that goes beyond τ_e (Holder et al. 2003, Kaplinghat et al. 2003). The above studies have shown that two different reionization histories could produce the same value of τ_e but could predict different shapes of a polarization "bump," which could be distinguished at high significance in *Planck* data. In particular, a long period of partial ionization extending out to high redshift would shift power toward smaller angular scales. Heinrich et al. (2017) and Miranda et al. (2017) have recently fit parametric reionization

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

models to the public 2015 *Planck* Low Frequency Instrument polarization data. Interestingly, they concluded that this analysis favors reionization histories with somewhat elevated τ_e values ($\tau_e \approx 0.08$, compared to the value $\tau_e \approx 0.05$ obtained by assuming prescribed, sharp ionization histories; Planck Collab. 2018) and a shape that mimics a tail of partial ionization extending to high redshift (with 10–20% ionization at $z = 15$–20) but a relatively sudden full reionization at $z = 6$–7. A similar feature, however, could also be produced by a high-z tail of partial ionization from early stars (Ahn et al. 2012). Future, more sensitive CMB polarization experiments (aiming to detect signatures of primordial B-modes from inflation), as well as the analysis of *Planck*'s High Frequency Instrument data, can provide better measurements of any high-z preionization, e.g., the Stage-4 ground-based CMB experiment (CMB-S4; Abazajian et al. 2019) and LiteBIRD [Lite (Light) satellite for the studies of B-mode polarization and Inflation from cosmic background Radiation Detection; Matsumura et al. 2014, Hazumi et al. 2019].

The ionization and thermal history of the high-redshift IGM can also be probed by radio observations of the redshifted 21-cm line of neutral hydrogen (Furlanetto et al. 2006, Pritchard & Loeb 2012). These observations can break degeneracies between contributions to preionization by softer (UV) radiation from stars and by harder (X-ray) radiation from accreting BHs. The latter is expected to produce a much smoother spatial morphology than the "swiss-cheese" ionization structure produced by UV photons (Zhang et al. 2007, Mesinger et al. 2013). Interferometers such as Hydrogen Epoch of Reionization Arrays (HERA) are designed to measure spatial fluctuations in the emission/absorption signal, whereas global experiments such as EDGES [Experiment to Detect the Global EoR (epoch of reionization) Signature] seek to measure the signal averaged over the entire sky. One of the main predicted features of the global 21-cm signal is an upturn in brightness temperature corresponding to early X-ray heating from accreting BHs or stellar remnants. Tanaka et al. (2016) have argued that if Pop III stars serve as the seeds of the first SMBHs, this should result in early X-ray heating of the IGM observable in the 21-cm signal at $z > 20$. If this signal is not observed, it implies that SMBHs formed later with massive seeds, SMBHs only occur in a small fraction of galaxies at high redshift, or accreting BH seeds emit significantly fewer or softer X-rays than would be expected based on low-redshift observations (the impact of a softer X-ray spectrum is computed in Fialkov et al. 2014).

Recently, the EDGES experiment published the first detection of the global signal (Bowman et al. 2018), which shows a very strong absorption feature at $z \approx 20$ followed by rapid heating at $z \approx 16$. The large depth of this feature is puzzling and, to date, has no physically compelling interpretation without exotic assumptions. However, if additional observations confirm that the feature (even if with a more realistic, reduced amplitude) is cosmological and not a residual from foreground or instrumental modeling, this would contain interesting new information on the formation of the first SMBHs.

Another indirect method to probe the first SMBHs is via the unresolved X-ray background. Using again the Soltan–Paczyński argument (Soltan 1982), it is possible to put constraints on the total amount of high-redshift BH accretion. Dijkstra et al. (2004) have shown that this strongly limits the number density of faint, undetected BHs at high redshift; in particular, these BHs cannot contribute significantly to reionization without overproducing the unresolved X-ray background from *Chandra*. In a similar analysis, Salvaterra et al. (2012) derived limits of $\rho_\bullet \lesssim 0.7 \times 10^4\ \mathrm{M_\odot\ Mpc^{-3}}$ on the BH mass density at $z > 5$. However, these constraints can be significantly weakened by varying assumptions about the spectral properties of the BH emission (Cappelluti et al. 2017). Due to the high abundance of BHs, seed models based on Pop III stars that accrete efficiently are the most likely to be ruled out by X-ray background measurements (Ricarte & Natarajan 2018). In a related analysis, one can look for X-ray emission from BHs by stacking the X-ray observations of known optically detected galaxies. Treister et al. (2013) have stacked *Chandra* data for galaxies

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

found in the Hubble Deep Fields and have derived upper limits on the average X-ray luminosities of BHs in $z = 6$–8 galaxies. These limits can be translated to a mass density with the same Soltan–Paczynski approach as above and yield $\rho_\bullet \lesssim 10^3\ M_\odot\ \mathrm{Mpc}^{-3}$, implying (with further assumptions about the frequency of occurrence and spectrum of BHs in individual galaxies) an upper limit of $\sim 3 \times 10^6\ M_\odot$ on the mass of the nuclear SMBH in a typical $z \approx 6$ galaxy.

7.3. Fossil Evidence in the Local Universe

Local observations of IMBHs also have the potential to constrain high-redshift BHs [see the review by Greene et al. (2020) in this volume]. Due to their relatively quiet merger and star-formation histories, dwarf galaxies may share similarities with galaxies in the high-redshift Universe and retain information on BH seeding mechanisms. Theoretical predictions based on Monte Carlo halo merger trees suggest that the BH occupation fraction in these galaxies depends on the high-z seed properties (Volonteri et al. 2008, van Wassenhove et al. 2010). Because Pop III stars are expected to be much more abundant than massive seeds, models with stellar-mass BHs formed from these stars tend to have higher occupation fractions. However, if these seeds have not grown much beyond their initial masses, they will be very difficult to observe. It has also been suggested that the low-z $M_\bullet$–σ relation could be sensitive to SMBH seeding (Volonteri & Natarajan 2009), with massive seeds leading to a flatter low-mass end (i.e., BHs with masses above the relation found at large σ). However, subsequent work suggests that this relation depends more strongly on the accretion model than the seeding prescription (Ricarte & Natarajan 2018).

BHs have been discovered in several local dwarf galaxies, many with masses below $10^6\ M_\odot$ (Greene & Ho 2007, Barth et al. 2008, Reines et al. 2013, Moran et al. 2014, Baldassare et al. 2018, Chilingarian et al. 2018). The lowest-mass confirmed BH resides in the galaxy RGG 118 and has an estimated mass of $\sim 5 \times 10^4\ M_\odot$ (Baldassare et al. 2015). Additionally, there are a number of dynamical BH candidates in globular clusters with similar masses (e.g., Lützgendorf et al. 2013), but their lack of X-ray or radio emission signatures makes it difficult to confirm their existence (Strader et al. 2012, Wrobel et al. 2015). For a comprehensive review on local observations of IMBHs, see Mezcua (2017).

Although these observations are intriguing, the current BH occupation fraction is not well-enough constrained to draw strong conclusions about the SMBH seeding mechanism. The fact that IMBHs have only been found with $M_\bullet > 10^4\ M_\odot$ along with the observed flattening of the $M_\bullet$–σ relation at low masses (e.g., see figure 10 in Mezcua 2017) may hint at the importance of massive seeds. However, this inference should be taken with caution. A population of lower-mass IMBHs may have escaped detection to date, or stellar-mass seeds formed early on in dwarf galaxies could have grown substantially over time.

One potential challenge in studying massive seeds with BHs in dwarf galaxies is that their abundance is expected to be exceedingly low. As discussed above, if massive seed formation relies on full photodissociation of H_2, the required LW flux is very high, resulting in an abundance perhaps as low as $\approx 1\ \mathrm{Gpc}^{-3}$, just matching that of the bright $z \approx 6$ quasars (Dijkstra et al. 2014). In this case, we could not expect to find such a seed in the local Universe. Other formation channels can create many more massive BHs, but, as discussed above, the overall number density is still quite low. Taking the synchronized halo pairs scenario (Visbal et al. 2014b) as an example, and assuming that massive seed formation is shut off by metal pollution at relatively high redshift, the number density of $\approx 10^{-3}\ \mathrm{Mpc}^{-3}$ would represent roughly the total abundance of massive seeds created throughout the entire Universe. For this density, the fraction of dwarf galaxies containing massive seed BH fossils would be at most $\sim 10^{-2}$. Thus, if such massive BHs are found in dwarf galaxies, and their formation can be securely placed to high z (e.g., from the old age of the host's stellar

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

population), then this will imply that either the abundance of these seeds is much higher than expected or some lower-mass BHs grew efficiently. A recent study did find much larger occupation fractions: Most halos somewhat above the ACH limit, followed in simulations by Bellovary et al. (2019), were found to form massive seed BHs at $z \sim 15$–20. These simulations included treatments of H_2 chemistry and metal enrichment but did not resolve the small-scale collapse dynamics or the history of gas (and stars) in the earlier stages of these halos (including minihalos). Higher-resolution simulations are required to assess whether massive BHs could indeed be produced so often.

8. CONCLUSIONS

In an ideal world, this review would describe a robust prediction for the evolving BH mass function—together with the corresponding LF—as a function of redshift. The state of the field is far from this goal. However, there are well-developed ideas for the formation of BHs that range from stellar masses ($\sim$10–100 $M_\odot$) to BHs as massive as $\sim 10^6$ $M_\odot$ in early protogalaxies. This review has therefore focused mostly on describing these ideas.

In analogy with the stellar IMF, there is an IMF of BHs describing the distribution of BH masses at their birth. It is nearly certain that this IMF has a strong peak at stellar masses ($\sim$10–100 $M_\odot$), because the first generation of metal-poor stars were likely massive, and often (perhaps in the majority of cases) left behind BH remnants. It is also very likely that BHs more massive than these were formed by the processes described in this review and summarized in **Figure 3**. It is often useful to contrast these massive seeds with the light stellar-mass BHs for the purposes of illustrating extreme scenarios. However, the high-z BH IMF more likely covers the full range of $10 \lesssim M_\bullet/M_\odot \lesssim 10^6$.

The most massive BHs with $M_\bullet \approx 10^{5–6}$ $M_\odot$ require very special conditions—pristine metal-free gas with a suppressed H_2 abundance in relatively large ACHs with masses of more than about several $\times 10^7$ $M_\odot$—producing rapid, monolithic infall of 10^6 $M_\odot$ of gas. This will be realized only in rare, highly biased regions of the Universe that contain ACHs that are exposed to unusually intense Lyman–Werner radiations, have unusually and rapid assembly histories, and reside in regions with unusually high baryonic streaming velocity—or some combination of these three. However, we expect that IMBHs ($\sim 10^2$–10^4 $M_\odot$) should form in larger abundance in regions that are biased but fail to fully meet the above conditions for massive seed BH formation. For example, IMBHs could arise in overdense regions where H_2 cooling is suppressed but not fully disabled (so that an embryonic SMS does not grow to the massive BH regime) or where gas condensing in ACHs had some modest prior star formation and metal enrichment (so that the gas undergoes some fragmentation rather than accreting onto a sole central protostar). We generally expect that criteria to form increasingly massive BHs are increasingly harder to realize, and therefore BH IMFs will span the full range of $\sim$10–10^6 $M_\odot$, monotonically declining with mass in this range.

The subsequent growth of these seed BHs due to accretion and mergers will determine the evolution of their mass function over cosmic time. Some lucky BHs will find themselves in the dense cores of growing galaxies and will be able to accrete efficiently, but the majority, especially at early times, will fail to grow. This is because both gas and BHs can be relatively easily ejected from the shallow potential wells of the first microgalaxies via radiative processes and SN explosions and via merger-induced gravitational recoil, respectively. The BHs hosted by the most massive and most rapidly growing host halos will have the best chance to grow efficiently with a high duty cycle and to evolve into the quasars observed at $z \sim 6$–7. However, feedback from the SMBHs and accompanying stars is likely important in regulating their growth even in large ($10^{10–12}$ $M_\odot$) galaxies. This feedback, in the form of SNe, as well as radiation and mechanical energy from strong

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

outflows driven by the SMBH itself, requires simulations spanning the full dynamical range from where this radiation and outflow is generated to galactic scales, where feedback operates over cosmological timescales. Understanding the details of this is the challenging frontier in numerical simulations.

Finally, the admittedly poor state of our ab initio understanding of the early BH population is an opportunity for observations. Measuring the LF of quasars in optical and X-ray bands, one to two orders of magnitude below the present limits, should be feasible and should yield strong constraints on assembly models, especially at $z \gtrsim 10$. These constraints will likely suffer from degeneracies between seeding and growth, but these degeneracies should be lifted by the direct GW detections of merging BHs in the $\sim 10^4$–10^6 $M_\odot$ range by LISA and by combinations of indirect probes of early BHs via their imprint on the cosmic 21-cm signal and on large-scale CMB polarization anisotropies.

In order to probe the specific seed models, it will be necessary to detect BHs with masses of $\lesssim 10^5$ $M_\odot$ at redshifts $z \gtrsim 10$, because the newly born BHs will likely lose the memory of their birth by the time they grow well above $\sim 10^5$ $M_\odot$ and are incorporated into more massive host galaxies. Although challenging, it should be feasible to detect 10^5 $M_\odot$ BHs in the future in ultra-deep observations with X-ray telescopes such as *Lynx*, and somewhat more massive BHs in the optical/IR with JWST and next-generation, 30m-class optical/IR telescopes.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Yacine Ali-Haïmoud, Eduardo Bañados, Jillian Bellovary, Mitch Begelman, Xiaohui Fan, Katie Freese, Jenny Greene, Melanie Habouzit, Tilman Hartwig, Takashi Hosokawa, Yoshiki Matsuoka, Priya Natarajan, Kazu Omukai, Fabio Pacucci, Jim Stone, Hiromichi Tagawa, Fabio Vito, Marta Volonteri, Feige Wang, Jemma Wolcott-Green, Tyrone Woods, Jinyi Yang, and Naoki Yoshida for numerous useful comments on an earlier version of this manuscript, and Kyungjin Ahn, Tiziana Di Matteo, and Yoshiki Matsuoka for permission to adapt figures from their published work. K.I. acknowledges recent support from the National Science Foundation of China (11721303, 11991052, 11950410493) and the National Key R&D Program of China (2016YFA0400702). Z.H. acknowledges support from NASA through grants NNX15AB19G and NNX17AL82G and from the National Science Foundation through grant 1715661.

LITERATURE CITED

Abazajian K, Addison G, Adshead P, et al. 2019. arXiv:1907.04473
Abbott BP, Abbott R, Abbott TD, et al. 2019. *Phys. Rev. X* 9:031040
Abel T, Bryan GL, Norman ML. 2002. *Science* 295:93–98
Abramowicz MA, Czerny B, Lasota JP, Szuszkiewicz E. 1988. *Ap. J.* 332:646–58
Agarwal B, Cullen F, Khochfar S, et al. 2017. *MNRAS* 468:L82–86
Agarwal B, Davis AJ, Khochfar S, Natarajan P, Dunlop JS. 2013. *MNRAS* 432:3438–44
Agarwal B, Khochfar S. 2015. *MNRAS* 446:160–68
Agarwal B, Khochfar S, Johnson JL, et al. 2012. *MNRAS* 425:2854–71
Ahn K, Iliev IT, Shapiro PR, et al. 2012. *Ap. J. Lett.* 756:L16
Ahn K, Shapiro PR, Iliev IT, Mellema G, Pen UL. 2009. *Ap. J.* 695:1430–45

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Akiyama M, He W, Ikeda H, et al. 2018. *Publ. Astron. Soc. Jpn.* 70:S34
Alexander T, Natarajan P. 2014. *Science* 345:1330–33
Ali-Haïmoud Y, Kamionkowski M. 2017. *Phys. Rev. D* 95:043534
Ali-Haïmoud Y, Kovetz ED, Kamionkowski M. 2017. *Phys. Rev. D* 96:123523
Alvarez MA, Wise JH, Abel T. 2009. *Ap. J. Lett.* 701:L133–37
Amaro-Seoane P, Audley H, Babak S, et al. 2017. arXiv:1702.00786
Ardaneh K, Luo Y, Shlosman I, et al. 2018. *MNRAS* 479:2277–93
Bañados E, Venemans BP, Decarli R, et al. 2016. *Ap. J. Suppl.* 227:11
Bañados E, Venemans BP, Mazzucchelli C, et al. 2018. *Nature* 553:473–76
Bañados E, Venemans BP, Morganson E, et al. 2015. *Ap. J.* 804:118
Balberg S, Shapiro SL. 2002. *Phys. Rev. Lett.* 88:101301
Balbus SA, Hawley JF. 1991. *Ap. J.* 376:214–22
Baldassare VF, Geha M, Greene J. 2018. *Ap. J.* 868:152
Baldassare VF, Reines AE, Gallo E, Greene JE. 2015. *Ap. J. Lett.* 809:L14
Balmaverde B, Gilli R, Mignoli M, et al. 2017. *Astron. Astrophys.* 606:A23
Barkana R, Haiman Z, Ostriker JP. 2001. *Ap. J.* 558:482–96
Barnes J, Efstathiou G. 1987. *Ap. J.* 319:575–600
Barnes JE, Hernquist LE. 1991. *Ap. J. Lett.* 370:L65–68
Barrow KSS, Aykutalp A, Wise JH. 2018. *Nat. Astron.* 2:987–94
Barth AJ, Greene JE, Ho LC. 2008. *Astron. J.* 136:1179–200
Bean R, Magueijo J. 2002. *Phys. Rev. D* 66:063505
Becerra F, Greif TH, Springel V, Hernquist LE. 2015. *MNRAS* 446:2380–93
Becerra F, Marinacci F, Inayoshi K, Bromm V, Hernquist LE. 2018. *Ap. J.* 857:138
Becker RH, White RL, Helfand DJ. 1995. *Ap. J.* 450:559–77
Begelman MC. 1978. *MNRAS* 184:53–67
Begelman MC. 1979. *MNRAS* 187:237–51
Begelman MC, Rossi EM, Armitage PJ. 2008. *MNRAS* 387:1649–59
Begelman MC, Shlosman I. 2009. *Ap. J. Lett.* 702:L5–8
Begelman MC, Volonteri M, Rees MJ. 2006. *MNRAS* 370:289–98
Bellovary JM, Cleary CE, Munshi F, et al. 2019. *MNRAS* 482:2913–23
Belotsky KM, Dokuchaev VI, Eroshenko YN, et al. 2019. *Eur. Phys. J.* 79(3):246
Bertoldi F, Cox P, Neri R, et al. 2003. *Astron. Astrophys.* 409:L47–50
Bird S, Cholis I, Muñoz JB, et al. 2016. *Phys. Rev. Lett.* 116:201301
Birnboim Y, Dekel A. 2003. *MNRAS* 345:349–64
Blandford RD, Begelman MC. 2004. *MNRAS* 349:68–86
Boekholt TCN, Schleicher DRG, Fellhauer M, et al. 2018. *MNRAS* 476:366–80
Bolton JS, Haehnelt MG. 2007. *MNRAS* 374:493–514
Bondi H. 1952. *MNRAS* 112:195
Bowman JD, Rogers AEE, Monsalve RA, Mozdzen TJ, Mahesh N. 2018. *Nature* 555:67–70
Bromm V, Coppi PS, Larson RB. 2002. *Ap. J.* 564:23–51
Bromm V, Loeb A. 2003. *Ap. J.* 596:34–46
Bromm V, Yoshida N. 2011. *Annu. Rev. Astron. Astrophys.* 49:373–407
Buchner J, Bauer FE. 2017. *MNRAS* 465(4):4348–62
Bullock JS, Boylan-Kolchin M. 2017. *Annu. Rev. Astron. Astrophys.* 55:343–87
Bullock JS, Dekel A, Kolatt TS, et al. 2001. *Ap. J.* 555:240–57
Campanelli M, Lousto CO, Zlochower Y, Merritt D. 2007. *Phys. Rev. Lett.* 98:231102
Cappelluti N, Li Y, Ricarte A, et al. 2017. *Ap. J.* 837:19
Carilli CL, Walter F. 2013. *Annu. Rev. Astron. Astrophys.* 51:105–61
Carnall AC, Shanks T, Chehade B, et al. 2015. *MNRAS* 451:L16–20
Carr B, Kühnel F, Sandstad M. 2016. *Phys. Rev. D* 94:083504
Carr B, Silk J. 2018. *MNRAS* 478:3756–75

Carr BJ. 2006. In *59th Yamada Conference on Inflating Horizon of Particle Astrophysics and Cosmology, Tokyo, Japan*, June 20–24, 2005, pp. 129–50. Tokyo, Jpn.: Univ. Acad. Press
Carr BJ, Bond JR, Arnett WD. 1984. *Ap. J.* 277:445–69
Cen R, Haiman Z. 2000. *Ap. J. Lett.* 542:L75–78
Chambers KC, Magnier EA, Metcalfe N, et al. 2016. arXiv:1612.05560
Chandrasekhar S. 1964. *Ap. J.* 140:417–33
Chehade B, Carnall AC, Shanks T, et al. 2018. *MNRAS* 478:1649–59
Chen KJ, Heger A, Woosley S, et al. 2014. *Ap. J.* 790:162
Chilingarian IV, Katkov IY, Zolotukhin IY, et al. 2018. *Ap. J.* 863:1
Choi JH, Shlosman I, Begelman MC. 2013. *Ap. J.* 774:149
Choi JH, Shlosman I, Begelman MC. 2015. *MNRAS* 450:4411–23
Chon S, Hirano S, Hosokawa T, Yoshida N. 2016. *Ap. J.* 832:134
Choquette J, Cline JM, Cornell JM. 2019. *J. Cosmol. Astropart. Phys.* July 2019(07):036
Christodoulou DM, Shlosman I, Tohline JE. 1995. *Ap. J.* 443:551–62
Ciardi B, Ferrara A, Abel T. 2000. *Ap. J.* 533:594–600
Ciotti L, Ostriker JP. 2001. *Ap. J.* 551:131–52
Clark PC, Glover SCO, Klessen RS. 2008. *Ap. J.* 672:757–64
Coleman J. 2018. Paper presented at *39th International Conference on High Energy Physics (ICHEP2018), Seoul, Korea, July 4–11*. arXiv:1812.00482
Collin S, Kawaguchi T. 2004. *Astron. Astrophys.* 426:797–808
Comerford JM, Haiman Z, Schaye J. 2002. *Ap. J.* 580:63–72
Davies FB, Hannawi JF, Eilers A-C. 2019. *Ap. J. Lett.* 884:L19
Davies MB, Miller MC, Bellovary JM. 2011. *Ap. J. Lett.* 740:L42
Davis AJ, Natarajan P. 2009. *MNRAS* 393:1498–502
Dayal P, Choudhury TR, Bromm V, Pacucci F. 2017. *Ap. J.* 836:16
De Rosa G, Decarli R, Walter F, et al. 2011. *Ap. J.* 739:56
De Rosa G, Venemans BP, Decarli R, et al. 2014. *Ap. J.* 790:145
Decarli R, Dotti M, Bañados E, et al. 2019. *Ap. J.* 880:157
Decarli R, Walter F, Venemans BP, et al. 2017. *Nature* 545:457–61
Decarli R, Walter F, Venemans BP, et al. 2018. *Ap. J.* 854:97
Dekel A, Birnboim Y. 2006. *MNRAS* 368:2–20
Dekel A, Sari R, Ceverino D. 2009. *Ap. J.* 703:785–801
DES Collab., Abbott T, Aldering G, et al. 2005. arXiv:astro-ph/0510346
Devecchi B, Volonteri M. 2009. *Ap. J.* 694:302–13
Dey A, Schlegel DJ, Lang D, et al. 2019. *Astron. J.* 157:168
Di Matteo T, Colberg J, Springel V, Hernquist L, Sijacki D. 2008. *Ap. J.* 676:33–53
Di Matteo T, Khandai N, DeGraf C, et al. 2012. *Ap. J. Lett.* 745:L29
Dijkstra M, Ferrara A, Mesinger A. 2014. *MNRAS* 442:2036–47
Dijkstra M, Gronke M, Sobral D. 2016a. *Ap. J.* 823:74
Dijkstra M, Haiman Z, Loeb A. 2004. *Ap. J.* 613:646–54
Dijkstra M, Haiman Z, Mesinger A, Wyithe JSB. 2008. *MNRAS* 391:1961–72
Dijkstra M, Sethi S, Loeb A. 2016b. *Ap. J.* 820:10
Dolgov AD. 2018. *Int. J. Mod. Phys. A* 33:1844029
Dolgov AD, Kawasaki M, Kevlishvili N. 2009. *Nucl. Phys. B* 807:229–50
Draine BT, Bertoldi F. 1996. *Ap. J.* 468:269–89
Du P, Hu C, Lu KX, et al. 2014. *Ap. J.* 782:45
Dubois Y, Pichon C, Devriendt J, et al. 2013. *MNRAS* 428:2885–900
Dubois Y, Pichon C, Haehnelt M, et al. 2012. *MNRAS* 423:3616–30
Dubois Y, Volonteri M, Silk J, et al. 2015. *MNRAS* 452:1502–18
Dunn G, Bellovary J, Holley-Bockelmann K, Christensen C, Quinn T. 2018. *Ap. J.* 861:39
Dye S, Lawrence A, Read MA, et al. 2018. *MNRAS* 473:5113–25
Edge A, Sutherland W, Kuijken K, et al. 2013. *Messenger* 154:32–34

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Eilers A-C, Davies FB, Hennawi JF, et al. 2017. *Ap. J.* 840:24
Eilers A-C, Hennawi JF, Davies FB, et al. 2018. *Ap. J.* 867:30
Eisenstein DJ, Loeb A. 1995. *Ap. J.* 443:11–17
El-Neaj YA, Alpigiani C, Amairi-Pyky S, et al. 2020. *EPJ Quantum Technol.* 7:6
Escala A. 2007. *Ap. J.* 671:1264–71
Fabian AC. 2012. *Annu. Rev. Astron. Astrophys.* 50:455–89
Fan X. 2006. *New Astron. Rev.* 50:665–71
Fan X, Barth A, Banados E, et al. 2019a. arXiv:1903.04078
Fan X, Narayanan VK, Lupton RH, et al. 2001. *Astron. J.* 122:2833–49
Fan X, Strauss MA, Schneider DP, et al. 2003. *Astron. J.* 125:1649–59
Fan X, Wang F, Yang J, et al. 2019b. *Ap. J. Lett.* 870:L11
Fan X, White RL, Davis M, et al. 2000. *Astron. J.* 120:1167–74
Fanidakis N, Macciò AV, Baugh CM, Lacey CG, Frenk CS. 2013. *MNRAS* 436:315–26
Fernandez R, Bryan GL, Haiman Z, Li M. 2014. *MNRAS* 439:3798–807
Ferrara A, Haardt F, Salvaterra R. 2013. *MNRAS* 434:2600–5
Fialkov A, Barkana R, Tseliakhovich D, Hirata CM. 2012. *MNRAS* 424:1335–45
Fialkov A, Barkana R, Visbal E. 2014. *Nature* 506:197–99
Field GB, Somerville WB, Dressler K. 1966. *Annu. Rev. Astron. Astrophys.* 4:207–44
Fragile PC, Olejar A, Anninos P. 2014. *Ap. J.* 796:22
Freese K, Gondolo P, Sellwood JA, Spolyar D. 2009. *Ap. J.* 693:1563–69
Freese K, Rindler-Daller T, Spolyar D, Valluri M. 2016. *Rep. Prog. Phys.* 79:066902
Freitag M, Gürkan MA, Rasio FA. 2006. *MNRAS* 368:141–61
Fricke KJ. 1973. *Ap. J.* 183:941–58
Furlanetto SR, Oh SP, Briggs FH. 2006. *Phys. Rep.* 433:181–301
Furlanetto SR, Stoever SJ. 2010. *MNRAS* 404:1869–78
Gabor JM, Capelo PR, Volonteri M, et al. 2016. *Astron. Astrophys.* 592:A62
Gallerani S, Fan X, Maiolino R, Pacucci F. 2017. *Publ. Astron. Soc. Aust.* 34:e022
Genzel R, Eisenhauer F, Gillessen S. 2010. *Rev. Mod. Phys.* 82:3121–95
Glover SCO. 2015a. *MNRAS* 451:2082–96
Glover SCO. 2015b. *MNRAS* 453:2901–18
Goodman J. 2003. *MNRAS* 339:937–48
Goodman J, Tan JC. 2004. *Ap. J.* 608:108–18
Goulding AD, Greene JE, Bezanson R, et al. 2018. *Publ. Astron. Soc. Jpn.* 70(SP1):S37
Greene JE, Ho LC. 2007. *Ap. J.* 670:92–104
Greene JE, Strader J, Ho LC. 2020. *Annu. Rev. Astron. Astrophys.* 58:257–312
Greif TH. 2015. *Comput. Astrophys. Cosmol.* 2:3
Greif TH, Bromm V, Clark PC, et al. 2012. *MNRAS* 424:399–415
Greif TH, White SDM, Klessen RS, Springel V. 2011. *Ap. J.* 736:147
Gültekin K, Miller MC, Hamilton DP. 2004. *Ap. J.* 616:221–30
Habouzit M, Volonteri M, Dubois Y. 2017. *MNRAS* 468:3935–48
Habouzit M, Volonteri M, Latif M, Dubois Y, Peirani S. 2016. *MNRAS* 463:529–40
Habouzit M, Volonteri M, Somerviller RS, et al. 2019. *MNRAS* 489:1206–29
Haehnelt MG, Natarajan P, Rees MJ. 1998. *MNRAS* 300:817–27
Haemmerlé L, Meynet G, Mayer L, et al. 2019. *Astron. Astrophys.* 632:L2
Haemmerlé L, Woods TE, Klessen RS, Heger A, Whalen DJ. 2018. *MNRAS* 474:2757–73
Haiman Z. 2004. *Ap. J.* 613:36–40
Haiman Z. 2013. In *The First Galaxies*, ed. T Wiklind, B Mobasher, V Bromm, *Ap. Space Sci. Libr.* 396:293–341. Dordrecht, Neth.: Springer-Verlag
Haiman Z, Abel T, Rees MJ. 2000. *Ap. J.* 534:11–24
Haiman Z, Kocsis B, Menou K. 2009. *Ap. J.* 700:1952–69
Haiman Z, Loeb A. 2001. *Ap. J.* 552:459–63

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Haiman Z, Quataert E. 2004. In *Supermassive Black Holes in the Distant Universe*, ed. AJ Barger, *Astrophys. Space Sci. Libr.* 308:147–85. Dordrecht, Neth.: Kluwer Acad.
Haiman Z, Rees MJ, Loeb A. 1996a. *Ap. J.* 467:522–31
Haiman Z, Rees MJ, Loeb A. 1997. *Ap. J.* 476:458–63
Haiman Z, Thoul AA, Loeb A. 1996b. *Ap. J.* 464:523–38
Hanawa T, Matsumoto T. 2000. *Publ. Astron. Soc. Jpn.* 52:241–47
Hartwig T, Agarwal B, Regan JA. 2018a. *MNRAS* 479:L23–27
Hartwig T, Bromm V, Loeb A. 2018b. *MNRAS* 479:2202–13
Hartwig T, Glover SCO, Klessen RS, Latif MA, Volonteri M. 2015. *MNRAS* 452:1233–44
Hawking S. 1971. *MNRAS* 152:75–78
Hayashi C. 1961. *Publ. Astron. Soc. Jpn.* 13:450–52
Hazumi M, Ade PAR, Akiba Y, et al. 2019. *J. Low Temp. Phys.* 194:443–52
Heckman TM, Best PN. 2014. *Annu. Rev. Astron. Astrophys.* 52:589–660
Heger A, Woosley SE. 2002. *Ap. J.* 567:532–43
Heinrich CH, Miranda V, Hu W. 2017. *Phys. Rev. D* 95:023513
Herrmann F, Hinder I, Shoemaker D, Laguna P, Matzner RA. 2007. *Ap. J.* 661:430–36
Hirano S, Hosokawa T, Yoshida N, Kuiper R. 2017. *Science* 357:1375–78
Hirano S, Hosokawa T, Yoshida N, Omukai K, Yorke HW. 2015. *MNRAS* 448:568–87
Hirano S, Hosokawa T, Yoshida N, et al. 2014. *Ap. J.* 781:60
Hirano S, Yoshida N, Sakurai Y, Fujii MS. 2018. *Ap. J.* 855:17
Hirasawa T. 1969. *Prog. Theor. Phys.* 42:523–43
Holder GP, Haiman Z, Kaplinghat M, Knox L. 2003. *Ap. J.* 595:13–18
Hopkins PF, Hernquist L, Cox TJ, Kereš D. 2008. *Ap. J. Suppl.* 175:356–89
Hopkins PF, Quataert E. 2010. *MNRAS* 407:1529–64
Hopkins PF, Quataert E. 2011. *MNRAS* 415:1027–50
Hosokawa T, Omukai K, Yorke HW. 2012. *Ap. J.* 756:93
Hosokawa T, Omukai K, Yoshida N, Yorke HW. 2011. *Science* 334:1250–53
Hosokawa T, Yorke HW, Inayoshi K, Omukai K, Yoshida N. 2013. *Ap. J.* 778:178
Igumenshchev IV, Abramowicz MA. 1999. *MNRAS* 303:309–20
Igumenshchev IV, Narayan R, Abramowicz MA. 2003. *Ap. J.* 592:1042–59
Ilie C, Freese K, Valluri M, Iliev IT, Shapiro PR. 2012. *MNRAS* 422:2164–86
Inayoshi K, Haiman Z. 2014. *MNRAS* 445:1549–57
Inayoshi K, Haiman Z, Ostriker JP. 2016. *MNRAS* 459:3738–55
Inayoshi K, Hosokawa T, Omukai K. 2013. *MNRAS* 431:3036–44
Inayoshi K, Li M, Haiman Z. 2018. *MNRAS* 479:4017–27
Inayoshi K, Omukai K. 2011. *MNRAS* 416:2748–59
Inayoshi K, Omukai K. 2012. *MNRAS* 422:2539–46
Inayoshi K, Omukai K, Tasker E. 2014. *MNRAS* 445:L109–13
Inayoshi K, Tanaka TL. 2015. *MNRAS* 450:4350–63
Inayoshi K, Visbal E, Kashiyama K. 2015. *MNRAS* 453:1692–700
Inutsuka Si, Miyama SM. 1997. *Ap. J.* 480:681–93
Izumi T, Onoue M, Shirakata H, et al. 2018. *Publ. Astron. Soc. Jpn.* 70:36
Jeon M, Pawlik AH, Greif TH, et al. 2012. *Ap. J.* 754:34
Jiang L, Fan X, Annis J, et al. 2008. *Astron. J.* 135:1057–66
Jiang L, McGreer ID, Fan X, et al. 2016. *Ap. J.* 833:222
Jiang YF, Stone JM, Davis SW. 2014. *Ap. J.* 796:106
Jiang YF, Stone JM, Davis SW. 2019. *Ap. J.* 880:67
Johnson JL, Bromm V. 2007. *MNRAS* 374:1557–68
Johnson JL, Dalla VC, Khochfar S. 2013. *MNRAS* 428:1857–72
Johnson JL, Dijkstra M. 2017. *Astron. Astrophys.* 601:A138
Johnson JL, Haardt F. 2016. *Publ. Astron. Soc. Aust.* 33:e007
Johnson JL, Khochfar S, Greif TH, Durier F. 2011. *MNRAS* 410:919–33

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Johnson JL, Whalen DJ, Agarwal B, Paardekooper JP, Khochfar S. 2014. *MNRAS* 445:686–93
Jungman G, Kamionkowski M, Griest K. 1996. *Phys. Rep.* 267:195–373
Kaplinghat M, Chu M, Haiman Z, et al. 2003. *Ap. J.* 583:24–32
Kasen D, Woosley SE, Heger A. 2011. *Ap. J.* 734:102
Kashikawa N, Ishizaki Y, Willott CJ, et al. 2015. *Ap. J.* 798:28
Kashiyama K, Inayoshi K. 2016. *Ap. J.* 826:80
Katz H, Sijacki D, Haehnelt MG. 2015. *MNRAS* 451:2352–69
Kauffmann G, Haehnelt M. 2000. *MNRAS* 311:576–88
Kawashima T, Ohsuga K, Mineshige S, et al. 2012. *Ap. J.* 752:18
Keeton CR, Kuhlen M, Haiman Z. 2005. *Ap. J.* 621:559–73
Kereš D, Katz N, Weinberg DH, Davé R. 2005. *MNRAS* 363:2–28
Khandai N, Feng Y, DeGraf C, Di Matteo T, Croft RAC. 2012. *MNRAS* 423:2397–406
Khlopov MY. 2010. Res. *Astron. Astrophys.* 10:495–528
Kim S, Stiavelli M, Trenti M, et al. 2009. *Ap. J.* 695:809–17
King A, Lasota JP, Kluźniak W. 2017. *MNRAS* 468:L59–62
King AR, Davies MB, Ward MJ, Fabbiano G, Elvis M. 2001. *Ap. J. Lett.* 552:L109–12
Kitayama T, Yoshida N. 2005. *Ap. J.* 630:675–88
Kitayama T, Yoshida N, Susa H, Umemura M. 2004. *Ap. J.* 613:631–45
Klein A, Barausse E, Sesana A, et al. 2016. *Phys. Rev. D* 93:024003
Klessen RS, Glover SCO. 2016. *Saas-Fee Adv. Course* 43:85
Koda J, Shapiro PR. 2011. *MNRAS* 415:1125–37
Koppitz M, Pollney D, Reisswig C, et al. 2007. *Phys. Rev. Lett.* 99:041102
Koptelova E, Hwang CY, Yu PC, Chen WP, Guo JK. 2017. *Sci. Rep.* 7:41617
Kormendy J, Ho LC. 2013. *Annu. Rev. Astron. Astrophys.* 51:511–653
Koushiappas SM, Bullock JS, Dekel A. 2004. *MNRAS* 354:292–304
Krumholz MR, Klein RI, McKee CF, Offner SSR, Cunningham AJ. 2009. *Science* 323:754–57
Kulkarni M, Visbal E, Bryan GL. 2019. *Ap. J.* 882:178
Kunze KE, Komatsu E. 2015. *J. Cosmol. Astropart. Phys.* 2015:027
Lai D. 2000. *Ap. J.* 540:946–61
Larson RB. 1969. *MNRAS* 145:271–95
Larson RB. 1985. *MNRAS* 214:379–98
Larson RB. 2003. *Rep. Prog. Phys.* 66:1651–97
Latif M, Schleicher D. 2018. *Formation of the First Black Holes*. Singapore: World Sci. Publ. Co.
Latif MA, Bovino S, Grassi T, Schleicher DRG, Spaans M. 2015. *MNRAS* 446:3163–77
Latif MA, Bovino S, Van Borm C, et al. 2014a. *MNRAS* 443:1979–87
Latif MA, Ferrara A. 2016. *Publ. Astron. Soc. Aust.* 33:e051
Latif MA, Khochfar S. 2019. *MNRAS* 490:2706–16
Latif MA, Schleicher DRG. 2015. *Astron. Astrophys.* 578:A118
Latif MA, Schleicher DRG, Hartwig T. 2016. *MNRAS* 458:233–41
Latif MA, Schleicher DRG, Schmidt W. 2014b. *MNRAS* 440:1551–61
Latif MA, Schleicher DRG, Schmidt W, Niemeyer J. 2013. *MNRAS* 433:1607–18
Latif MA, Volonteri M, Wise JH. 2018. *MNRAS* 476:5016–25
Lauer TR, Faber SM, Ajhar EA, Grillmair CJ, Scowen PA. 1998. *Astron. J.* 116:2263–86
Lauer TR, Faber SM, Richstone D, et al. 2007. *Ap. J.* 662:808–34
Lawrence A, Warren SJ, Almaini O, et al. 2007. *MNRAS* 379:1599–617
Levine R, Gnedin NY, Hamilton AJS, Kravtsov AV. 2008. *Ap. J.* 678:154–67
Li Y, Hernquist L, Robertson B, et al. 2007. *Ap. J.* 665:187–208
Lin MK, Krumholz MR, Kratter KM. 2011. *MNRAS* 416:580–90
Lodato G, Natarajan P. 2006. *MNRAS* 371:1813–23
Lodato G, Natarajan P. 2007. *MNRAS* 377:L64–68
Luo Y, Ardaneh K, Shlosman I, et al. 2018. *MNRAS* 476:3523–39
Lupi A, Haardt F, Dotti M, et al. 2016. *MNRAS* 456:2993–3003

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Lupi A, Volonteri M, Decarli R, et al. 2019. *MNRAS* 488:4004–22
Lützgendorf N, Kissler-Patig M, Gebhardt K, et al. 2013. *Astron. Astrophys.* 552:A49
Lynden-Bell D. 1969. *Nature* 223:690–94
Lynden-Bell D, Kalnajs AJ. 1972. *MNRAS* 157:1–30
Machacek ME, Bryan GL, Abel T. 2001. *Ap. J.* 548:509–21
Machida MN, Doi K. 2013. *MNRAS* 435:3283–305
Madau P, Haardt F, Dotti M. 2014. *Ap. J. Lett.* 784:L38
Madau P, Haardt F, Rees MJ. 1999. *Ap. J.* 514:648–59
Madau P, Rees MJ. 2001. *Ap. J. Lett.* 551:L27–30
Madau P, Rees MJ, Volonteri M, Haardt F, Oh SP. 2004. *Ap. J.* 604:484–94
Maki H, Susa H. 2007. *Publ. Astron. Soc. Jpn.* 59:787–97
Matsuda T, Satō H, Takeda H. 1969. *Prog. Theor. Phys.* 42:219–33
Matsumoto T, Nakauchi D, Ioka K, Heger A, Nakamura T. 2015. *Ap. J.* 810:64
Matsumoto T, Nakauchi D, Ioka K, Nakamura T. 2016. *Ap. J.* 823:83
Matsumura T, Akiba Y, Borrill J, et al. 2014. *J. Low Temp. Phys.* 176:733–40
Matsuoka Y, Iwasawa K, Onoue M, et al. 2018a. *Ap. J. Suppl.* 237:5
Matsuoka Y, Iwasawa K, Onoue M, et al. 2019a. *Ap. J.* 883:183
Matsuoka Y, Onoue M, Kashikawa N, et al. 2016. *Ap. J.* 828:26
Matsuoka Y, Onoue M, Kashikawa N, et al. 2018b. *Publ. Astron. Soc. Jpn.* 70:S35
Matsuoka Y, Onoue M, Kashikawa N, et al. 2019b. *Ap. J. Lett.* 872:L2
Matsuoka Y, Strauss MA, Kashikawa N, et al. 2018c. *Ap. J.* 869:150
Mayer L, Bonoli S. 2019. *Rep. Prog. Phys.* 82:016901
Mayer L, Fiacconi D, Bonoli S, et al. 2015. *Ap. J.* 810:51
Mayer L, Kazantzidis S, Escala A, Callegari S. 2010. *Nature* 466:1082–84
Mayer L, Kazantzidis S, Madau P, et al. 2007. *Science* 316:1874–77
Mazzucchelli C, Bañados E, Venemans BP, et al. 2017. *Ap. J.* 849:91
McGreer ID, Becker RH, Helfand DJ, White RL. 2006. *Ap. J.* 652:157–62
McGreer ID, Fan X, Jiang L, Cai Z. 2018. *Astron. J.* 155:131
McGreer ID, Fan X, Strauss MA, et al. 2014. *Astron. J.* 148:73
McKee CF, Tan JC. 2008. *Ap. J.* 681:771–97
McKinney JC, Dai L, Avara MJ. 2015. *MNRAS* 454:L6–10
McKinney JC, Tchekhovskoy A, Sadowski A, Narayan R. 2014. *MNRAS* 441:3177–208
McMahon RG, Banerji M, Gonzalez E, et al. 2013. *Messenger* 154:35–37
Mellema G, Iliev IT, Alvarez MA, Shapiro PR. 2006. *New Astron.* 11:374–95
Merritt D, Szell A. 2006. *Ap. J.* 648:890–99
Mesinger A, Bryan GL, Haiman Z. 2006. *Ap. J.* 648:835–51
Mesinger A, Ferrara A, Spiegel DS. 2013. *MNRAS* 431:621–37
Mesinger A, Haiman Z. 2004. *Ap. J. Lett.* 611:L69–72
Meyer RA, Bosman SEI, Ellis RS. 2019. *MNRAS* 487:3305–23
Mezcua M. 2017. *Int. J. Mod. Phys. D* 26:1730021
Miller MC. 2000. *Ap. J.* 544:43–48
Milosavljević M, Bromm V, Couch SM, Oh SP. 2009a. *Ap. J.* 698:766–80
Milosavljević M, Couch SM, Bromm V. 2009b. *Ap. J. Lett.* 696:L146–49
Mineshige S, Kawaguchi T, Takeuchi M, Hayashida K. 2000. *Publ. Astron. Soc. Jpn.* 52:499–508
Miranda V, Lidz A, Heinrich CH, Hu W. 2017. *MNRAS* 467:4050–56
Mo HJ, Mao S, White SDM. 1998. *MNRAS* 295:319–36
Montero PJ, Janka HT, Müller E. 2012. *Ap. J.* 749:37
Moran EC, Shahinyan K, Sugarman HR, Vélez DO, Eracleous M. 2014. *Astron. J.* 148:136
Morganson E, De Rosa G, Decarli R, et al. 2012. *Astron. J.* 143:142
Mortlock DJ, Warren SJ, Venemans BP, et al. 2011. *Nature* 474:616–19
Nakauchi D, Inayoshi K, Omukai K. 2014. *MNRAS* 442:2667–79
Nanni R, Vignali C, Gilli R, Moretti A, Brand t WN. 2017. *Astron. Astrophys.* 603:A128

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Narayan R, Igumenshchev IV, Abramowicz MA. 2003. *Publ. Astron. Soc. Jpn.* 55:L69–72
Narita S, Hayashi C, Miyama SM. 1984. *Prog. Theor. Phys.* 72:1118–36
Natarajan P, Pacucci F, Ferrara A, et al. 2017. *Ap. J.* 838:117
Neeleman M, Bañados E, Walter F, et al. 2019. *Ap. J.* 882:10
Nguyen DD, Seth AC, Neumayer N, et al. 2018. *Ap. J.* 858:118
Novak M, Bañados E, Decarli R, et al. 2019. *Ap. J.* 881:63
Oh SP, Haiman Z. 2002. *Ap. J.* 569:558–72
Oh SP, Haiman Z, Rees MJ. 2001. *Ap. J.* 553:73–77
Ohsuga K, Mori M, Nakamoto T, Mineshige S. 2005. *Ap. J.* 628:368–81
Okuda T. 2002. *Publ. Astron. Soc. Jpn.* 54:253–66
Omukai K. 2001. *Ap. J.* 546:635–51
Omukai K, Nishi R. 1998. *Ap. J.* 508:141–50
Omukai K, Nishi R. 1999. *Ap. J.* 518:64–68
Omukai K, Palla F. 2001. *Ap. J. Lett.* 561:L55–58
Omukai K, Palla F. 2003. *Ap. J.* 589:677–87
Omukai K, Schneider R, Haiman Z. 2008. *Ap. J.* 686:801–14
Onoue M, Kashikawa N, Matsuoka Y, et al. 2019. *Ap. J.* 880:77
O'Shea BW, Norman ML. 2008. *Ap. J.* 673:14–33
Ostriker JP. 2000. *Phys. Rev. Lett.* 84:5258–60
Ostriker JP, Peebles PJE. 1973. *Ap. J.* 186:467–80
Ota K, Venemans BP, Taniguchi Y, et al. 2018. *Ap. J.* 856:109
Pacucci F, Ferrara A, Grazian A, et al. 2016. *MNRAS* 459:1432–39
Pacucci F, Ferrara A, Volonteri M, Dubus G. 2015a. *MNRAS* 454:3771–77
Pacucci F, Loeb A. 2019. *Ap. J. Lett.* 870:L12
Pacucci F, Loeb A, Mezcua M, Martín-Navarro I. 2018. *Ap. J. Lett.* 864:L6
Pacucci F, Mesinger A, Haiman Z. 2013. *MNRAS* 435:L53–57
Pacucci F, Volonteri M, Ferrara A. 2015b. *MNRAS* 452:1922–33
Padoan P, Nordlund Å. 2002. *Ap. J.* 576:870–79
Page DN, Hawking SW. 1976. *Ap. J.* 206:1–7
Palla F, Salpeter EE, Stahler SW. 1983. *Ap. J.* 271:632–41
Paoletti D, Chluba J, Finelli F, Rubiño-Martín JA. 2019. *MNRAS* 484:185–95
Park K, Ricotti M. 2011. *Ap. J.* 739:2
Park K, Ricotti M. 2012. *Ap. J.* 747:9
Peebles PJE, Dicke RH. 1968. *Ap. J.* 154:891–908
Penston MV. 1969. *MNRAS* 144:425–48
Peterson BM. 2006. In *Physics of Active Galactic Nuclei at all Scales*, ed. D Alloin, R Johnson, P Lira, *Lect. Notes Phys.* 693:77–100. Berlin: Springer-Verlag
Peterson BM, Ferrarese L, Gilbert KM, et al. 2004. *Ap. J.* 613:682–99
Pezzulli E, Valiante R, Schneider R. 2016. *MNRAS* 458:3047–59
Pfister H, Volonteri M, Dubois Y, Dotti M, Colpi M. 2019. *MNRAS* 486:101–11
Planck Collab., Ade PAR, Aghanim N, et al. 2016. *Astron. Astrophys.* 594:A19
Planck Collab., Aghanim N, Akrami Y, et al. 2018. *Astron. Astrophys.* Submitted. arXiv:1807.06209
Pollack J, Spergel DN, Steinhardt PJ. 2015. *Ap. J.* 804:131
Pons E, McMahon RG, Simcoe RA, et al. 2019. *MNRAS* 484:5142–54
Portegies Zwart SF, Baumgardt H, Hut P, Makino J, McMillan SLW. 2004. *Nature* 428:724–26
Poulin V, Serpico PD, Calore F, Clesse S, Kohri K. 2017. *Phys. Rev. D* 96:083524
Poutanen J, Lipunova G, Fabrika S, Butkevich AG, Abolmasov P. 2007. *MNRAS* 377:1187–94
Prieto J, Escala A. 2016. *MNRAS* 460:4018–37
Prieto J, Jimenez R, Haiman Z, González RE. 2015. *MNRAS* 452:784–802
Pritchard JR, Loeb A. 2012. *Rep. Prog. Phys.* 75:086901
Proga D, Begelman MC. 2003. *Ap. J.* 592:767–81
Quataert E, Gruzinov A. 2000. *Ap. J.* 539:809–14

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Randall SW, Markevitch M, Clowe D, Gonzalez AH, Bradač M. 2008. *Ap. J.* 679:1173–80
Reed SL, Banerji M, Becker GD, et al. 2019. *MNRAS* 487:1874–85
Reed SL, McMahon RG, Banerji M, et al. 2015. *MNRAS* 454:3952–61
Reed SL, McMahon RG, Martini P, et al. 2017. *MNRAS* 468:4702–18
Rees MJ. 1984. *Annu. Rev. Astron. Astrophys.* 22:471–506
Rees MJ, Ostriker JP. 1977. *MNRAS* 179:541–59
Regan JA, Downes TP. 2018. *MNRAS* 475:4636–47
Regan JA, Downes TP, Volonteri M, et al. 2019. *MNRAS* 486:3892–906
Regan JA, Haehnelt MG. 2009. *MNRAS* 396:343–53
Regan JA, Johansson PH, Haehnelt MG. 2014a. *MNRAS* 439:1160–75
Regan JA, Johansson PH, Wise JH. 2014b. *Ap. J.* 795:137
Regan JA, Johansson PH, Wise JH. 2016. *MNRAS* 459:3377–94
Regan JA, Visbal E, Wise JH, et al. 2017. *Nat. Astron.* 1:0075
Reines AE, Greene JE, Geha M. 2013. *Ap. J.* 775:116
Reinoso B, Schleicher DRG, Fellhauer M, Klessen RS, Boekholt TCN. 2018. *Astron. Astrophys.* 614:A14
Reisswig C, Ott CD, Abdikamalov E, et al. 2013. *Phys. Rev. Lett.* 111:151101
Ricarte A, Natarajan P. 2018. *MNRAS* 481:3278–92
Richards GT, Haiman Z, Pindor B, et al. 2006. *Astron. J.* 131:49–54
Ricotti M, Gnedin NY, Shull JM. 2001. *Ap. J.* 560:580–91
Ricotti M, Gnedin NY, Shull JM. 2002. *Ap. J.* 575:33–48
Ricotti M, Ostriker JP. 2004. *MNRAS* 352:547–62
Ricotti M, Ostriker JP, Gnedin NY. 2005. *MNRAS* 357:207–19
Ricotti M, Ostriker JP, Mack KJ. 2008. *Ap. J.* 680:829–45
Rindler-Daller T, Montgomery MH, Freese K, Winget DE, Paxton B. 2015. *Ap. J.* 799:210
Ritter JS, Safranek-Shrader C, Gnat O, Milosavljević M, Bromm V. 2012. *Ap. J.* 761:56
Ryu T, Tanaka TL, Perna R, Haiman Z. 2016. *MNRAS* 460:4122–34
Sądowski A. 2009. *Ap. J. Suppl.* 183:171–78
Sądowski A, Narayan R, Tchekhovskoy A, et al. 2015. *MNRAS* 447:49–71
Safarzadeh M, Loeb A. 2019. *Ap. J. Lett.* 877:L27
Saigo K, Hanawa T. 1998. *Ap. J.* 493:342–50
Saijo M, Hawke I. 2009. *Phys. Rev. D* 80:064001
Sakurai Y, Inayoshi K, Haiman Z. 2016a. *MNRAS* 461:4496–504
Sakurai Y, Vorobyov EI, Hosokawa T, et al. 2016b. *MNRAS* 459:1137–45
Sakurai Y, Yoshida N, Fujii MS, Hirano S. 2017. *MNRAS* 472:1677–84
Salvaterra R, Haardt F, Volonteri M, Moretti A. 2012. *Astron. Astrophys.* 545:L6
Saslaw WC, Zipoy D. 1967. *Nature* 216:976–78
Sato S, Kawamura S, Ando M, et al. 2017. *J. Phys. Conf. Ser.* 840:012010
Schauer ATP, Regan J, Glover SCO, Klessen RS. 2017. *MNRAS* 471:4878–84
Schleicher DRG, Banerjee R, Sur S, et al. 2010a. *Astron. Astrophys.* 522:A115
Schleicher DRG, Galli D, Glover SCO, et al. 2009. *Ap. J.* 703:1096–106
Schleicher DRG, Spaans M, Glover SCO. 2010b. *Ap. J. Lett.* 712:L69–72
Schmidt M. 1963. *Nature* 197:1040
Sesana A, Gair J, Berti E, Volonteri M. 2011. *Phys. Rev. D* 83:044036
Sesana A, Volonteri M, Haardt F. 2007. *MNRAS* 377:1711–16
Sethi S, Haiman Z, Pandey K. 2010. *Ap. J.* 721:615–21
Sethi SK, Nath BB, Subramanian K. 2008. *MNRAS* 387:1589–96
Shang C, Bryan GL, Haiman Z. 2010. *MNRAS* 402:1249–62
Shankar F, Salucci P, Granato GL, De Zotti G, Danese L. 2004. *MNRAS* 354:1020–30
Shanks T, Metcalfe N, Chehade B, et al. 2015. *MNRAS* 451:4238–52
Shapiro SL, Teukolsky SA. 1983. *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects*. New York: Wiley-Intersci.
Shen Y, Wu J, Jiang L, et al. 2019. *Ap. J.* 873:35

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Shibata M, Sekiguchi Y, Uchida H, Umeda H. 2016. *Phys. Rev. D* 94:021501
Shibata M, Shapiro SL. 2002. *Ap. J. Lett.* 572:L39–43
Shibuya T, Ouchi M, Harikane Y, et al. 2018. *Publ. Astron. Soc. Jpn.* 70:S15
Shimasaku K, Izumi T. 2019. *Ap. J. Lett.* 872:L29
Shlosman I, Begelman MC, Frank J. 1990. *Nature* 345:679–86
Shlosman I, Frank J, Begelman MC. 1989. *Nature* 338:45–47
Sijacki D, Springel V, Di Matteo T, Hernquist L. 2007. *MNRAS* 380:877–900
Sijacki D, Springel V, Haehnelt MG. 2009. *MNRAS* 400:100–22
Silk J. 1977. *Ap. J.* 211:638–48
Skrutskie MF, Cutri RM, Stiening R, et al. 2006. *Astron. J.* 131:1163–83
Smidt J, Whalen DJ, Chatzopoulos E, et al. 2015. *Ap. J.* 805:44
Smidt J, Whalen DJ, Johnson JL, Surace M, Li H. 2018. *Ap. J.* 865:126
Smith A, Becerra F, Bromm V, Hernquist L. 2017. *MNRAS* 472:205–16
Smith BD, Regan JA, Downes TP, et al. 2018. *MNRAS* 480:3762–73
Sobral D, Matthee J, Darvish B, et al. 2015. *Ap. J.* 808:139
Soltan A. 1982. *MNRAS* 200:115–22
Somerville RS, Hopkins PF, Cox TJ, Robertson BE, Hernquist L. 2008. *MNRAS* 391:481–506
Spergel DN, Steinhardt PJ. 2000. *Phys. Rev. Lett.* 84:3760–63
Spolyar D, Freese K, Gondolo P. 2008. *Phys. Rev. Lett.* 100:051101
Springel V, White SDM, Jenkins A, et al. 2005. *Nature* 435:629–36
Stacy A, Bromm V. 2007. *MNRAS* 382:229–38
Stacy A, Bromm V, Lee AT. 2016. *MNRAS* 462:1307–28
Stacy A, Bromm V, Loeb A. 2011. *Ap. J. Lett.* 730:L1
Stacy A, Greif TH, Bromm V. 2012. *MNRAS* 422:290–309
Stahler SW, Palla F, Salpeter EE. 1986. *Ap. J.* 302:590–605
Stecher TP, Williams DA. 1967. *Ap. J. Lett.* 149:L29–30
Stern D, Kirkpatrick JD, Allen LE, et al. 2007. *Ap. J.* 663:677–85
Stone JM, Pringle JE, Begelman MC. 1999. *MNRAS* 310:1002–16
Stone NC, Küpper AHW, Ostriker JP. 2017. *MNRAS* 467:4180–99
Strader J, Chomiuk L, Maccarone TJ, et al. 2012. *Ap. J. Lett.* 750:L27
Strader J, Seth AC, Forbes DA, et al. 2013. *Ap. J. Lett.* 775:L6
Subramanian K. 2016. *Rep. Prog. Phys.* 79:076901
Sugimura K, Hosokawa T, Yajima H, Inayoshi K, Omukai K. 2018. *MNRAS* 478:3961–75
Sugimura K, Hosokawa T, Yajima H, Omukai K. 2017a. *MNRAS* 469:62–79
Sugimura K, Mizuno Y, Matsumoto T, Omukai K. 2017b. *MNRAS* 469:4022–33
Sugimura K, Omukai K, Inoue AK. 2014. *MNRAS* 445:544–53
Sun L, Ruiz M, Shapiro SL. 2018. *Phys. Rev. D* 98:103008
Surace M, Whalen DJ, Hartwig T, et al. 2018. *Ap. J. Lett.* 869:L39
Susa H. 2007. *Ap. J.* 659:908–17
Susa H, Hasegawa K, Tominaga N. 2014. *Ap. J.* 792:32
Tagawa H, Haiman Z, Kocsis B. 2019. *Ap. J.* 892:36
Tagawa H, Umemura M, Gouda N. 2016. *MNRAS* 462:3812–22
Takahashi HR, Ohsuga K. 2015. *Publ. Astron. Soc. Jpn.* 67:60
Takahashi K, Yoshida T, Umeda H. 2018. *Ap. J.* 857:111
Takeo E, Inayoshi K, Ohsuga K, Takahashi HR, Mineshige S. 2018. *MNRAS* 476:673–82
Takeo E, Inayoshi K, Ohsuga K, Takahashi HR, Mineshige S. 2019. *MNRAS* 488:2689–700
Tanaka T, Haiman Z. 2009. *Ap. J.* 696:1798–822
Tanaka T, Perna R, Haiman Z. 2012. *MNRAS* 425:2974–87
Tanaka TL. 2014. *Class. Quantum Gravity* 31:244005
Tanaka TL, Li M. 2014. *MNRAS* 439:1092–100
Tanaka TL, O'Leary RM, Perna R. 2016. *MNRAS* 455:2619–26
Tang JJ, Goto T, Ohyama Y, et al. 2017. *MNRAS* 466:4568–72

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Tchekhovskoy A, Narayan R, McKinney JC. 2011. *MNRAS* 418:L79–83
Tegmark M, Silk J, Rees MJ, et al. 1997. *Ap. J.* 474:1–12
Thompson TA, Quataert E, Murray N. 2005. *Ap. J.* 630:167–85
Toyouchi D, Hosokawa T, Sugimura K, Nakatani R, Kuiper R. 2019. *MNRAS* 483:2031–43
Trakhtenbrot B, Volonteri M, Natarajan P. 2017. *Ap. J. Lett.* 836:L1
Treister E, Schawinski K, Volonteri M, Natarajan P. 2013. *Ap. J.* 778:130
Tremmel M, Karcher M, Governato F, et al. 2017. *MNRAS* 470:1121–39
Truelove JK, Klein RI, McKee CF, et al. 1997. *Ap. J. Lett.* 489:L179–83
Tseliakhovich D, Hirata C. 2010. *Phys. Rev. D* 82:083520
Tumlinson J, Shull JM. 2000. *Ap. J.* 528:L65–68
Turk MJ, Oishi JS, Abel T, Bryan GL. 2012. *Ap. J.* 745:154
Turner EL. 1991. *Astron. J.* 101:5–17
Ueda Y, Akiyama M, Ohta K, Miyaji T. 2003. *Ap. J.* 598:886–908
Umeda H, Hosokawa T, Omukai K, Yoshida N. 2016. *Ap. J. Lett.* 830:L34
Utsumi Y, Goto T, Kashikawa N, et al. 2010. *Ap. J.* 721:1680–88
Valiante R, Agarwal B, Habouzit M, Pezzulli E. 2017. *Publ. Astron. Soc. Aust.* 34:e031
Valiante R, Schneider R, Volonteri M, Omukai K. 2016. *MNRAS* 457:3356–71
Valiante R, Schneider R, Zappacosta L, et al. 2018. *MNRAS* 476:407–20
Van Borm C, Bovino S, Latif MA, et al. 2014. *Astron. Astrophys.* 572:A22
Van Borm C, Spaans M. 2013. *Astron. Astrophys.* 553:L9
van den Bosch FC, Abel T, Croft RAC, Hernquist L, White SDM. 2002. *Ap. J.* 576:21–35
van Wassenhove S, Volonteri M, Walker MG, Gair JR. 2010. *MNRAS* 408:1139–46
Venemans BP, Bañados E, Decarli R, et al. 2015. *Ap. J.* 801:L11
Venemans BP, Findlay JR, Sutherland WJ, et al. 2013. *Ap. J.* 779:24
Venemans BP, Neeleman M, Walter F, et al. 2019. *Ap. J. Lett.* 874:L30
Venemans BP, Walter F, Decarli R, et al. 2017a. *Ap. J.* 837:146
Venemans BP, Walter F, Decarli R, et al. 2017b. *Ap. J. Lett.* 851:L8
Vestergaard M, Osmer PS. 2009. *Ap. J.* 699:800–16
Visbal E, Haiman Z. 2018. *Ap. J. Lett.* 865:L9
Visbal E, Haiman Z, Bryan GL. 2014a. *MNRAS* 442:L100–4
Visbal E, Haiman Z, Bryan GL. 2014b. *MNRAS* 445:1056–63
Visbal E, Haiman Z, Bryan GL. 2015. *MNRAS* 453:4456–66
Vito F, Brandt WN, Bauer FE, et al. 2019. *Astron. Astrophys.* 628:L6
Vito F, Brandt WN, Yang G, et al. 2018. *MNRAS* 473:2378–406
Volonteri M. 2010. *Astron. Astrophys.* 18:279–315
Volonteri M. 2012. *Science* 337:544–47
Volonteri M, Bellovary J. 2012. *Rep. Prog. Phys.* 75:124901
Volonteri M, Haardt F, Madau P. 2003. *Ap. J.* 582:559–73
Volonteri M, Lodato G, Natarajan P. 2008. *MNRAS* 383:1079–88
Volonteri M, Natarajan P. 2009. *MNRAS* 400:1911–18
Volonteri M, Rees MJ. 2006. *Ap. J.* 650:669–78
Volonteri M, Stark DP. 2011. *MNRAS* 417:2085–93
Walcher CJ, van der Marel RP, McLaughlin D, et al. 2005. *Ap. J.* 618:237–46
Walter F, Bertoldi F, Carilli C, et al. 2003. *Nature* 424:406–8
Walter F, Riechers D, Cox P, et al. 2009. *Nature* 457:699–701
Wang F, Wu XB, Fan X, et al. 2016a. *Ap. J.* 819:24
Wang F, Yang J, Fan X, et al. 2018. *Ap. J.* 869:L9
Wang F, Yang J, Fan X, et al. 2019. *Ap. J.* 884:30
Wang JM, Netzer H. 2003. *Astron. Astrophys.* 398:927–36
Wang R, Wagg J, Carilli CL, et al. 2013. *Ap. J.* 773:44
Wang R, Wu XB, Neri R, et al. 2016b. *Ap. J.* 830:53
Warren SJ, Hewett PC, Irwin MJ, McMahon RG, Bridgeland MT. 1987. *Nature* 325:131–33

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Watarai Ky, Fukue J, Takeuchi M, Mineshige S. 2000. *Publ. Astron. Soc. Jpn.* 52:133–41
Whalen D, Abel T, Norman ML. 2004. *Ap. J.* 610:14–22
Whalen D, van Veelen B, O'Shea BW, Norman ML. 2008. *Ap. J.* 682:49–67
Whalen DJ, Even W, Smidt J, et al. 2013. *Ap. J.* 778:17
Widrow LM. 2002. *Rev. Mod. Phys.* 74:775–823
Willott CJ, Albert L, Arzoumanian D, et al. 2010a. *Astron. J.* 140:546–60
Willott CJ, Bergeron J, Omont A. 2015. *Ap. J.* 801:123
Willott CJ, Bergeron J, Omont A. 2017. *Ap. J.* 850:108
Willott CJ, Delorme P, Omont A, et al. 2007. *Astron. J.* 134:2435–50
Willott CJ, Delorme P, Reylé C, et al. 2010b. *Astron. J.* 139:906–18
Winter LM, Mushotzky RF, Reynolds CS. 2006. *Ap. J.* 649:730–52
Wise JH, Abel T. 2007a. *Ap. J.* 665:899–910
Wise JH, Abel T. 2007b. *Ap. J.* 671:1559–67
Wise JH, Regan JA, O'Shea BW, et al. 2019. *Nature* 566:85–88
Wise JH, Turk MJ, Abel T. 2008. *Ap. J.* 682:745–57
Wise JH, Turk MJ, Norman ML, Abel T. 2012. *Ap. J.* 745:50
Wolcott-Green J, Haiman Z. 2012. *MNRAS* 425:L51–55
Wolcott-Green J, Haiman Z. 2019. *MNRAS* 484:2467–73
Wolcott-Green J, Haiman Z, Bryan GL. 2011. *MNRAS* 418:838–52
Wolcott-Green J, Haiman Z, Bryan GL. 2017. *MNRAS* 469:3329–36
Wolcott-Green J, Haiman Z, Bryan GL. 2020. *MNRAS*. Submitted. arXiv:2001.05498
Woods TE, Agarwal B, Bromm V, et al. 2019. *Publ. Astron. Soc. Aust.* 36:e027
Woods TE, Heger A, Whalen DJ, Haemmerlé L, Klessen RS. 2017. *Ap. J. Lett.* 842:L6
Wright EL, Eisenhardt PRM, Mainzer AK, et al. 2010. *Astron. J.* 140:1868–81
Wrobel JM, Nyland KE, Miller-Jones JCA. 2015. *Astron. J.* 150:120
Wu XB, Wang F, Fan X, et al. 2015. *Nature* 518:512–15
Wyithe JSB, Loeb A. 2002. *Ap. J.* 577:57–68
Wyithe JSB, Loeb A. 2004. *Nature* 577(6977):815–17
Yajima H, Khochfar S. 2016. *MNRAS* 457:2423–32
Yajima H, Ricotti M, Park K, Sugimura K. 2017. *Ap. J.* 846:3
Yang J, Wang F, Fan X, et al. 2019. *Astron. J.* 157:236
York DG, Adelman J, Anderson JE Jr., et al. 2000. *Astron. J.* 120:1579–87
Yorke HW, Sonnhalter C. 2002. *Ap. J.* 569:846–62
Yoshida N, Abel T, Hernquist L, Sugiyama N. 2003. *Ap. J.* 592:645–63
Yoshida N, Omukai K, Hernquist L. 2008. *Science* 321:669–71
Yoshida N, Omukai K, Hernquist L, Abel T. 2006. *Ap. J.* 652:6–25
Yu Q, Tremaine S. 2002. *MNRAS* 335:965–76
Zel'dovich YB, Novikov ID. 1967. *Sov. Astron.* 10:602–3
Zel'dovich YB, Novikov ID. 1971. *Relativistic Astrophysics. Vol. 1: Stars and Relativity*. Chicago: Univ. Chicago Press
Zhang J, Hui L, Haiman Z. 2007. *MNRAS* 375:324–36
Zucca A, Li Y, Pogosian L. 2017. *Phys. Rev. D* 95:063506
Zumalacárregui M, Seljak U. 2018. *Phys. Rev. Lett.* 121:141101

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Annual Review of Astronomy and Astrophysics

Volume 58, 2020

Contents

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.

Indexes

Errata

An online log of corrections to *Annual Review of Astronomy and Astrophysics* articles may be found at http://www.annualreviews.org/errata/astro

Annu. Rev. Astron. Astrophys. 2020.58:27-97. Downloaded from www.annualreviews.org
Access provided by Columbia University on 05/03/21. For personal use only.