The existence of a supermassive black hole (SMBH) at the centre of almost every galaxy provides strong evidence for a generic growth mechanism alongside that of the host galaxy (Kormendy & Ho 2013). Nuclear black holes typically comprise very much less than 1% of the available baryonic mass in the galaxy and are therefore gravitationally dominant only locally. But their efficient conversion of rest mass to radiation plays an important role in the evolution of galaxies as a powerful source of feedback that drives winds, suppresses star formation and contributes to the reionization of the universe. Understanding the physical processes leading to the formation and growth of SMBHs is crucial in unraveling fundamental questions about the galactic building blocks of the universe. While the largest-scale structures took several billion years to form, supermassive black holes up to $\sim 10^{10} M_\odot$ (quasars) are observed very early in the history of the universe, around a cosmological redshift of $z \approx 7$, at less than 5% of the current age of the universe (Fan et al. 2006; Mortlock et al. 2011; Wu et al. 2015). This presents a timing problem under conventional scenarios for the growth of these objects by accretion.

The rate of black hole growth via gas consumption is regulated by balancing the attractive gravitational force against the build-up of pressure from the radiation emitted by the infalling gas as it encounters viscous heating. For typical radiative efficiencies, the maximal growth rate – known as the Eddington limit – barely allows growth from stellar-mass black hole progenitors in the limited time available since the Big Bang. As an analogy, we consider the exponential growth of a bank account accruing compound interest. A lifetime of steady investments makes the burden of saving for retirement relatively straightforward and secure. However, if the growth has to occur in less than a decade then there is a timing crisis: the unfortunate individual must either save money at exorbitant rates, experience a miracle akin to winning the lottery – or delay their retirement.

**Timing Matters**

How and when did the first supermassive black holes form? Although there may not be a universal pathway, rapid seeding or growth is unavoidable within our current understanding of the most distant quasars. The first SMBHs were likely to have co-assembled...
within the first galaxies a few hundred million years after the Big Bang. Remarkably, primordial galaxies that could delay the onset of star formation seem to facilitate a mechanism to produce massive ($10^5-10^8 \, M_\odot$) black hole seeds, which form “in one go” from gas clouds with inefficient cooling. Such clouds thereby maintain high thermal pressure support, which in turn suppresses fragmentation and thus star formation. These direct collapse black holes (DCBHs) circumvent the timing crisis. DCBHs can form only early in cosmic history when a specific set of rare conditions is satisfied. At redshifts $z \gtrsim 10$, they are still beyond the reach of current telescopes. Progress will be made with next-generation observatories using wide-field surveys and follow-up deep observations of select objects. Ultimately, this will provide compelling evidence for or against formation scenarios involving direct-collapse, persistent hyper-Eddington accretion, or other more exotic possibilities (e.g. Dolgov & Silk 1993). In the sections below, we briefly review the historical development of ideas about the first SMBH seeds, the physics of their formation and radiative feedback, recent progress, and our outlook for the future. Figure 1 provides a schematic overview of the quasar-seed timing problem.

### Cosmological context

The standard cosmological model involves dark energy and cold dark matter ($\Lambda$CDM). Both ingredients remain mysterious in terms of their underlying physics. However, the $\Lambda$CDM model matches data across scales extending from large-scale structures to galaxy formation and evolution. Historically, supermassive black hole research predated the recent cosmological perspective. Summarizing these earlier insights, Rees (1984) discussed possible routes for runaway growth in active galactic nuclei. In isolated environments, possible formation pathways include: stellar-remnant black holes after vigorous gas accretion; dense star clusters in which runaway collisions trigger the formation of a SMBH; or a cluster of post-supernovae neutron stars or black holes coalescing in the dynamically unstable central core of the galaxy. In cosmology, SMBHs can grow over longer timescales via episodic galaxy mergers and accretion from streams of cold gas along filaments of the cosmic web (Mayer et al. 2010, 2015). Still, it was recognized that a gas cloud may conceivably bypass conventional star formation and yield a near-extremal Kerr black hole if cooling does not initiate fragmentation (Rees 1984).

Early simulations of collapsing primordial gas clouds showed that most of the gas does indeed fragment into dense stellar clumps which eventually virialize into a spheroidal galactic bulge (Loeb & Rasio 1994). The emergence of a SMBH would thus be forestalled. On the other hand, if a central seed black hole of mass $\gtrsim 10^8 \, M_\odot$ were to form during the collapse then it could quickly grow by steady accretion to a quasar-size black hole (Li et al. 2007). Otherwise dynamical instabilities in the inner region of the disc would inhibit the accretion required for such growth. The crucial bottleneck is the formation of the initial, million solar-mass, seed. Either way, realistic modeling of massive black hole inception is highly complex and ultimately relies on understanding both small- and large-scale phenomena. For example, low angular momentum configurations would help facilitate runaway collapse because the centrifugal barrier is significantly lower; in this context, quasar seeds could be a natural consequence of the initial collapse of regions with unusually small rotation (Eisenstein & Loeb 1995). Such low-spin cosmological perturbations might provide environments in which SMBH formation is an extreme manifestation of the $\Lambda$CDM model.

### Forming the first massive black holes

The main contenders for the earliest quasar seeds are DCBHs, super-Eddington accretion onto stellar remnant black holes, and runaway collisions in dense star clusters. Here we focus on DCBHs and highlight recent theoretical and observational evidence for this new class of black hole seeds; we discuss the alternative scenarios briefly.

In typical galactic environments black hole accretion is episodic because of self-regulating radiative feedback which yields accretion rates that are sub-Eddington when averaged over multiple duty cycles (Johnson & Bromm 2007; Milosavljević et al. 2009). However, maintaining super-Eddington accretion is possible when the black hole is embedded within sufficiently dense gas; this renders the radiation pressure less effective (e.g. Wyithe & Loeb 2012; Pacucci et al. 2015a). Based on one-dimensional radiation-hydrodynamics simulations, Inayoshi et al. (2016) find accretion rates exceeding $M_\bullet \gtrsim 10^5 \, L_\odot /c^2$ when the following condition is satisfied: $(n_\text{ms}/10^3 \, \text{cm}^{-3}) > (M_\bullet/10^8 \, M_\odot)^{-1} (T_\text{ms}/10^4 \, \text{K})^{1/2}$, where $n_\text{ms}$ and $T_\text{ms}$ are the density and temperature of the ambient gas. It remains an open question whether such growth rates are sustainable considering the violent assembly environments of the first galaxies where newly formed stars and supernovae blow away the surrounding gas. Other scenarios may also work, such as dense star clusters that undergo runaway collapse. With a ubiquitous supply of cold gas effectively trapping accretion radiation, a ~ $10^8 \, M_\odot$ black hole seed undergoing random motions through the cluster may initiate supra-exponential growth over a dynamical timescale (Alexander & Natarajan 2014).

On the other hand, forming a DCBH requires collapse without fragmentation, a cosmic “miracle” of sorts, that is naturally explained within the context of galaxy formation theory. The idea is that if a primordial gas cloud, devoid of any heavy chemical elements (“metals” in astronomical terminology), is bathed in a sea of ultraviolet radiation then it will be unable to cool and form stars (figure 2). Specifically, according to the Jeans criterion for triggering gravitational instability, high thermal pressure is required to prevent the gas cloud from fragmenting. In present-day star-forming clouds, line cooling by heavy elements and dust radiates away thermal energy that would otherwise provide stability.
But in the early universe only hydrogen and helium were available to cool the gas. Thus, a small abundance of molecular hydrogen ($H_2$) played a key role as the primary cooling agent below the $\sim 8000\,K$ accessible via atomic hydrogen line cooling. If strong non-ionizing Lyman-Werner radiation ( LW; with photon energies below $13.6\,eV$) from neighbouring galaxies photodissociates $H_2$ then the evolutionary track through density-temperature phase space is significantly altered (see Fig. 2). Only atomic cooling primarily through the Lyman-$\alpha$ ($Ly-\alpha$) line of hydrogen efficiently keeps the collapsing core below $\sim 10^4\,K$. The first idealized models of isothermal collapse including $H_2$ photodissociation still predicted that the cloud breaks up at late stages despite the higher temperature track (Omukai 2001). Gas collapse may be entirely suppressed in the minihaloes, with virial masses of $M_{\text{vir}} \lesssim 10^7\,M_{\odot}$, that are predicted to host the first stars, when the LW background flux is above a critical value. More massive host haloes, on the other hand, in general become self-shielding so that molecules can form again, eventually leading to star formation (Oh & Haiman 2002). However, Bromm & Loeb (2003), carrying out the first simulations of this collapse with cosmological initial conditions, recognized that under conditions of unusually strong LW irradiation, the infall would continue on its near-isothermal track. The resulting free-fall collapse of the atomically-cooling gas could then produce massive black holes directly. Subsequently, the DCBH model has received increased attention (e.g. Begelman et al. 2006; Regan & Haehnelt 2009; Choi et al. 2013, 2015). This body of work is broadly reviewed in Volonteri (2012), Haiman (2013), Loeb & Furlanetto (2013), Johnson & Haardt (2016), and Latif & Ferrara (2016).

Radiation from the first black holes

Active galactic nuclei are conspicuous manifestations of gas accretion onto supermassive black holes, with luminosities exceeding the total starlight of their host galaxies. The accretion disc radiates broadband emission from optical to X-ray wavelengths with a peak in the UV. The characteristic blackbody temperature for the Eddington luminosity near the event horizon is $T_{\text{Edd}} \approx 5 \times 10^8\,K (M_*/10^8\,M_{\odot})^{-1/4}$ (Rees 1984). DCBHs are born in gas-rich environments and may be self-shielding to ionizing photons. Still, the gas remains transparent to X-rays and the “Compton-thick” spectrum retains the non-thermal tail contributed by Bremsstrahlung radiation and magneto-hydrodynamical (MHD) processes (Pacucci et al. 2015b). In these environments, radiation pressure from the black hole along with concurrent star formation and nearby supernovae are likely to have had a substantial impact on the host galaxy (Jeon et al. 2012). Lower mass minihaloes with shallower gravitational potential wells ($M_{\text{vir}} \lesssim 10^7 - 10^8\,M_{\odot}$) would have been especially susceptible to radiative feedback, which potentially depleted the reservoir of gas needed to fuel black hole growth (Whalen et al. 2004; Wise et al. 2012). The viability of the DCBH mechanism also relies on the regulation of chemical feedback because molecular hydrogen and metal cooling induces fragmentation (Hartwig et al. 2016, and recall figure 2). Therefore, the eventual DCBH formation sites must remain free from star formation during collapse. Furthermore, the emission of UV and X-ray photons promotes $H_2$ formation and increases the critical LW flux needed to form DCBHs (Inayoshi & Tanaka 2015; Latif et al. 2015).

A particularly interesting source of feedback in DCBH environments is $Ly-\alpha$ photon trapping. In the vicinity of a newly collapsing primordial gas cloud, the surrounding neutral hydrogen is extremely optically thick to photons near the $Ly-\alpha$ resonant line. In one-dimensional simulations, a dense shell-like outflow structure forms in hydrodynamical response to the central source ionizing and heating of the gas. As illustrated in figure 3, the $Ly-\alpha$ photons are redshifted by the galactic outflow, producing a velocity offset that may result in considerably less scattering out of the line of sight by the intervening intergalactic medium (IGM). The complex nature of the $Ly-\alpha$ radiative transfer means that explorations of $Ly-\alpha$ feedback tend to focus on order-of-magnitude estimates based on idealized calculations (Oh & Haiman 2002; McKee & Tan 2008; Milosavljević et al. 2009). With the aid of post-processing Monte-Carlo radiative transfer (MCRT) methods, Dijkstra & Loeb (2008) found that multiple scattering within high HI column density shells is capable of enhancing the effective $Ly-\alpha$ force by one or two orders of magnitude. The first self-consistent $Ly-\alpha$ radiation hydrodynamics (RHD) simulations were performed by Smith et al. (2017), who coupled a MCRT code (corct; Smith et al. 2015) with spherically symmetric Lagrangian frame hydrodynamics including ionizing radiation, non-equilibrium chemistry and cooling, and self-
gravity. They found that Ly-α radiation pressure may have a significant dynamical impact on gas surrounding DCBHs, with Ly-α signatures characterized by larger velocity offsets than stellar counterparts if, in both cases, the Ly-α spectra are shaped by radiation-driven winds. Finally, it has recently been suggested that trapped Ly-α cooling radiation may enhance the formation of DCBHs by accounting for the photodetachment of H+ ions, precursors to H2, by Ly-α photons during collapse (7).

Observational evidence for DCBHs

The DCBH scenario has received increased popularity as theoretical predictions agreed with multiple lines of observational evidence. Although individual cases remain tenuous, there is good reason to believe SMBH progenitor candidates may be observationally confirmed with the capabilities of next-generation observatories. Recently, the luminous COSMOS redshift 7 (CR7) Ly-α emitter at z = 6.6 was confirmed to have exceptionally strong Ly-α (≥ 8 × 10^{42} erg s^{-1}) and possible He II 1640 Å (∼ 2 × 10^{43} erg s^{-1}) emission with no detection of metal lines from the UV to the near-infrared within instrumental sensitivity (Matthee et al. 2015; Sobral et al. 2015). As a result, several groups have considered the CR7 source in the context of a young primordial starburst or DCBH (Pallottini et al. 2015; Agarwal et al. 2016; Hartwig et al. 2016; Visbal et al. 2016; Dijkstra et al. 2016; Smidt et al. 2016). In Smith et al. (2016), we examined and reproduced several Ly-α signatures of the CR7 source under the Ly-α RHD framework discussed in the section above (see also figure 3). As shown in figure 4, the DCBH model reproduces the observed 160 km s^{-1} velocity offset between the Ly-α and He II line peaks, whereas the stellar model fails. We also found that Ly-α radiation pressure turns out to be dynamically important in the case of CR7.

However, more recently, 7 obtained deeper observations of CR7. The authors claim the new photometry cannot be reproduced by a DCBH spectral energy distribution (SED), suggesting instead that the broadband measurements may be contaminated by forbidden, doubly-ionized oxygen [O ii] emission lines. They propose that CR7 can be classified as a more standard low-mass, narrow-line AGN or a low-metallicity starburst getting the hard SED from massive stellar binaries. In contrast, Pacucci et al. (2017) argue that the new photometry is still consistent with the DCBH model. Either way, deep spectroscopy with future telescopes will be needed to discriminate convincingly between particular models (see also Agarwal et al. 2017). In the near future, other sources similar to CR7 may provide additional constraints on early galaxy and quasar formation. Indeed, Pacucci et al. (2016) identified two objects characterized by very red colours and robust X-ray detections in the IGM. The observed velocity offsets set of 160 km s^{-1} velocity offset between the Ly-α and He II line peaks, whereas the stellar model fails. We also found that Ly-α radiation pressure turns out to be dynamically important in the case of CR7.

Outlook for the future

Finally, we consider the prospects for unveiling the nature of supermassive black hole seeds with next-generation facilities. Continuing technological advances are going to allow us to probe the high-redshift universe in unprecedented detail. This includes the characterization of individual objects and integrated backgrounds based on observations in the radio, infrared, optical and X-ray wavelengths. Furthermore, space-based gravitational wave detectors will constrain SMBH merger models. Lastly, high-resolution simulations of black hole environments with multiscale physics will continue to refine our understanding of the formation and evolution of galactic black holes.

The James Webb Space Telescope (JWST) and giant segmented mirror telescopes1 may allow direct observations of the first galaxies and quasars. In particular, the Near-Infrared Camera (NIRCam) and Near-Infrared Spectrometer (NIRSpec) instruments on board the JWST will be capable of obtaining deep photometric and spectroscopic observations. The new data will yield a more complete census of galaxies from the first billion years of cosmic history, possibly including their redshifts, spectral energy distributions, star formation rates, metallicity, and emission line properties. As a reference, figure 5 illustrates the gas density from an ab initio cosmological simulation in which primordial gas undergoes the direct collapse to a black hole (Becerra et al. 2017). We show the galaxy at different scales to highlight the filamentary large-scale structure, gas distribution and opaque cloud within a sub-parsec region.

1 Telescopes with integral field spectrographs and adaptive optics imaging will include the Giant Magellan Telescope (GMT; www.gmto.org), Thirty Metre Telescope (TMT; www.tmt.org), and the European Extremely Large Telescope (E-ELT; www.eso.org/sci/facilities/eelt).
These filaments. The three panels show the cosmic web, host galaxy and the opaque gas cloud during SMBH seed formation at redshift $z \approx 11.6$ corresponding to 385 Myr after the Big Bang. For comparison, the $2.2' \times 2.2'$ field of view for the JWST is about 400 times larger in area than the left panel. The circles in the middle panel show the 0.032'' (solid) and 0.065'' (dashed) pixel resolution of the NIRCam instrument on board the JWST for the 2 $\mu$m and 4 $\mu$m channels, respectively. The central region where the DCBH forms is deeply embedded in the surrounding cloud (right panel) and remains spatially unresolved in this simulation. (Adapted from Becerra et al. (2017))

Several other complementary observatories will contribute to the emerging picture as well. For example, black holes often produce jets with strong radio emission, which may be observed at higher redshifts with the Atacama Large Millimeter Array (ALMA$^2$). Upcoming 21-cm cosmology experiments, such as the Square Kilometre Array (SKA$^3$), will map the distribution of neutral hydrogen over the course of early cosmic history through reionization, providing a better understanding of the contribution of high-redshift quasars to this process. Future observations will also provide better measurements of the cosmic infrared and X-ray backgrounds which exhibit a correlation that might be explained by unresolved massive black holes in faint galaxies. Eventually, many of these sources will be resolved and characterized as the proposed Lynx and Athena X-ray telescopes detect high-energy emission from distant SMBHs. Deep surveys of nearby dwarf galaxies might also reveal traces of the black hole seeding mechanism due to their relative isolation after formation. Finally, the planned Laser Interferometer Space Antenna (eLISA$^4$) promises to directly detect these massive black holes via mergers, extending gravitational wave astronomy from the stellar-mass events recently detected with the Laser Interferometer Gravitational-Wave Observatory (LIGO$^5$) to the DCBH mass range.

Simulations

The field of black hole research will benefit from the steady progress in computational algorithms and hardware, as high-resolution hydrodynamical simulations will provide additional insights into the process of SMBH formation. In particular, fully coupled radiation-hydrodynamics simulations will be crucial to elucidate the astrophysical phenomena responsible for the rapid growth of the first black holes. Pioneering simulations typically focus on specific aspects of larger questions in order to balance algorithmic complexity, resolution and computational feasibility. Eventually, it will be possible to apply more efficient and robust methods to problems with broader applicability. For example, incorporating effects based on accurate but traditionally expensive techniques, such as 3D Monte-Carlo radiative transfer, will be increasingly viable and worthwhile. The goal of these simulations is to connect what can be directly observed with what is ultimately powering these sources, however challenging this may be. A smoking-gun signature of an individual DCBH may be beyond the capabilities of next-generation telescopes, but the emergence of multiple independent lines of evidence might present a compelling picture in which massive black hole seeds bridge the gap in understanding the genesis of the first quasars.

REFERENCES

Agarwal B., Johnson J. L., Zakrisson E., Labbe I., van den Bosch F. C., Natarajan P., Khochfar S., 2016, MNRAS, 460, 4003
Agarwal B., Johnson J. L., Khochfar S., Pellegrini E., Rydberg C.-E., Klessen R. S., Oesch P., 2017, MNRAS, preprint (arXiv:1702.00407)
Alexander T., Natarajan P., 2014, Science, 345, 1330
Becerra F., Marinacci F., Inayoshi K., Bromm V., Hernquist L. E., 2017, preprint (arXiv:1702.03941)
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Bromm V., Loeb A., 2003, ApJ, 596, 34
Cappelluti N., et al., 2013, ApJ, 769, 68
Choi J.-H., Shlosman I., Begelman M. C., 2013, ApJ, 774, 149
Choi J.-H., Shlosman I., Begelman M. C., 2015, MNRAS, 450, 4411
Dijkstra M., Loeb A., 2008, MNRAS, 391, 457
Dijkstra M., Gronke M., Sobral D., 2016, ApJ, 823, 74
Dolgov A., Silk J., 1993, Physical Review D, 47, 4244
Eisenstein D. J., Loeb A., 1995, ApJ, 443, 11
Fan X., et al., 2006, AJ, 131, 1203
Haiman Z., 2013, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics

$^2$ www.almaobservatory.org
$^3$ skatelescope.org
$^4$ www.elisascience.org
$^5$ www.ligo.org
