The Race Between Stars and Quasars in Reionizing Cosmic Hydrogen

Abraham Loeb  
Astronomy Department, Harvard University, 60 Garden St., Cambridge, MA 02138, USA

The cosmological background of ionizing radiation has been dominated by quasars once the Universe aged by \( \sim 2 \) billion years. At earlier times (redshifts \( z \gtrsim 3 \)), the observed abundance of bright quasars declines sharply, implying that cosmic hydrogen was reionized by stars instead. Here, we explain the physical origin of the transition between the dominance of stars and quasars as a generic feature of structure formation in the concordance ΛCDM cosmology. At early times, the fraction of baryons in galaxies grows faster than the maximum (Eddington-limited) growth rate possible for quasars. As a result, quasars were not able to catch up with the rapid early growth of stellar mass in their host galaxies.

**Introduction.** Observations of the Lyα forest and the quasar luminosity function at redshifts \( z \lesssim 6 \), suggest that quasars dominated the production rate of hydrogen-ionizing photons only after the Universe has aged by 1-2 billion years [see Fig. 5 in Ref. 1]. The sharp decline in the observed comoving density of bright quasars at redshifts \( z \gtrsim 3 \) implies that supermassive black holes (BHs) could not have produced sufficient UV photons to reionize cosmic hydrogen by a redshift of \( z_{\text{reion}} = 10.9 \pm 1.4 \), as required by the WMAP data on the microwave background anisotropies [2]. It is therefore widely believed that stars have reionized cosmic hydrogen [1, 2]. Here we suggest a simple physical explanation for this phenomenological inference.

**Preliminaries.** Quasars are powered by the accretion of gas onto massive BHs [4, 5]. The remnant BHs they leave behind are observed in the nuclei of all bulged galaxies at the present time [6]. BH growth is accelerated during episodes of galaxy mergers, when cold gas is infused to galactic nuclei by tides; this association is predicted by computer simulations [7, 8, 9] and supported by observations [10]. The observed correlation between the mass of nuclear BHs and the depth of the gravitational potential well of their host galactic bulges (as inferred from the velocity dispersion of stars) [11], suggests that their growth was ultimately self-regulated [12, 13]. Merely \( \sim 5\% \) of the energy output from a bright quasar is sufficient to unbind the cold gas reservoir from its host galaxy [8, 13].

Bright quasars are inferred to be radiating close to their limiting Eddington [14] luminosity, \( L_E \). Estimates of their BH masses imply typical luminosities \( L \) in the range of \( \sim (0.1-1) \times L_E \) [see Fig. 6 in Ref. 15]. The Eddington limit is derived by equating the repulsive radiation force on an ionized gas element to the attractive gravitational force on it towards the BH [16],

\[
L_E \equiv \frac{4\pi c G m_p M}{\sigma_T} = 1.4 \times 10^{44} \text{ erg s}^{-1} \left( \frac{M}{10^9 M_\odot} \right),
\]

where \( \sigma_T \) is the Thomson cross-section for electron scattering and \( M \) is the BH mass. In a spherical geometry, the radiation force associated with \( L > L_E \) would generate an outflow and inhibit accretion onto the BH. Given an unlimited fuel reservoir, the growth rate of a BH would be regulated by the maximum luminosity that allows it to accrete gas, \( L_E \).

The luminosity of a quasar is related to the gas accretion rate \( \dot{M}_{\text{acc}} \) through \( L = \epsilon \dot{M}_{\text{acc}} c^2 \), where \( \epsilon \) is the efficiency for converting the rest mass of the accreting gas to radiation. At high accretion rates, the BH luminosity is expected to approach its limiting value \( L_{E, \text{lim}} \) as inferred for bright quasars [15]. Theoretical models imply that at \( L \gtrsim 0.5 L_E \), the accretion disk is puffed-up by radiation pressure [18, 19] and its geometry enters the quasi-spherical regime for which the Eddington limit is derived.

The growth rate of the BH mass, \( \dot{M} = (1 - \epsilon) \dot{M}_{\text{acc}} \) is related to the *Salpeter* [4] time,

\[
t_E = \frac{M}{\dot{M}} = 4 \times 10^8 \text{ yr} \left( \frac{\epsilon}{1 - \epsilon} \right) \left( \frac{L}{L_E} \right)^{-1}.
\]

For constant values of \( \epsilon \) and \( L/L_E \), the growth of the black hole from a seed mass \( \dot{M}_{\text{seed}} \) over time \( \Delta t \) is exponential, with \( M = \dot{M}_{\text{seed}} \exp(\Delta t/t_E) \). Interestingly, the growth time \( t_E \) has no explicit dependence on BH mass, and so it remains the same irrespective of whether multiple seeds grow in parallel or they coalesce to grow as a single BH.

At the high accretion rates of bright quasars, the cooling time of the gas is shorter than its accretion time [17] and a thin accretion disk with a high radiative efficiency [20] forms. The existence of such disks has been confirmed recently by microlensing observations [21]. The radiative efficiency of the gas is dictated by the inner boundary of the disk at the innermost stable circular orbit [22], from where the gas plunges into the BH and its viscous dissipation rate diminishes [19]. For a non-spinning BH, \( \epsilon = 5.7\% \), while for a maximally-spinning BH, \( \epsilon = 42\% \) [16, 22]. During the prodigious growth phase of quasars, the BH is expected to be spun-up quickly by the infalling gas, but the orientation of the angular momentum vector of the disk might vary considerably between different accretion episodes. Therefore, we adopt an intermediate value between these efficiencies in our fiducial numerical example. The fact that
the cosmic BH mass budget grew through a luminous accretion mode of a high radiative efficiency ($\epsilon \sim 10\%$) and not through a hidden mode of a low radiative efficiency, is demonstrated by comparing the radiation energy density produced by quasars to the BH mass density in the local Universe [23, 24]. A high radiative efficiency is also implied by the clustering and abundance statistics of quasars [25].

Stars form through fragmentation of cold gas in disks of galaxies [20]. The growth of the stellar mass budget occurs on the dynamical timescale of the host galactic region [2]. The most vigorous mode of star formation (starburst activity) is also realized in gas-rich galaxy mergers within which cold gas is concentrated into a compact region [7]. The fragmentation (Jeans) mass is lowered to the scale of stars only in environments that are denser by $\gtrsim 4$ orders of magnitude than the mean cosmic density $\bar{\rho}$; hence, the associated dynamical time is guaranteed to be shorter by $\gtrsim 2$ orders of magnitude than the age of the Universe $\sim (\bar{\rho})^{-1/2}$ at all redshifts. The undelayed growth in the stellar content of galaxies through merger-driven starbursts is ultimately limited by the global assembly rate of cold gas into galaxies. This rate is in turn dictated by the cosmological build-up of dark matter halos in which gas may cool.

The deposition of energy and momentum by stars or quasars into their gaseous environment may expel gas from the host galaxy and further regulate the growth of quasars into their gaseous environment may expel gas in which gas may cool.

The production rate of ionizing photons is proportional to the growth rate in the mass densities of BHs and stars per comoving volume in the Universe [14]. While the minimum growth time of BH mass is $t_E$, the shortest growth time of the stellar mass budget, $t_{\text{gal}}$, is dictated by the assembly rate of cold gas into galaxies. At late cosmic times, $t_E$ is much shorter than $t_{\text{gal}}$, leading to a feedback regulated mode of BH growth in which the supply of cold gas is a limiting factor [8, 12, 13]. Below we show that at early cosmic times the situation was reversed.

**Mass assembly rate of galaxies.** The minimum mass of galaxies in which BHs and stars form is dictated by cooling considerations. At the redshifts of interest here, the cooling time of the gas is shorter than its dynamical time for dark matter halos that have a virial temperature $T_{\text{vir}} \gtrsim 10^8 K$, above the threshold for atomic hydrogen cooling [28]. Galaxies above this threshold are believed to have hosted the bulk of the sources that have reionized the Universe. Early on, the minimum virial temperature of galaxies might have been reduced by an order of magnitude through molecular hydrogen cooling, but H$_2$ can be easily dissociated by UV photons [29]. After reionization, the minimum $T_{\text{vir}}$ is expected to have increased to $\sim 10^9 K$, owing to photo-ionization heating of the intergalactic medium from where galaxies are assembled [30].

The fraction of baryons available to make stars equals the mass fraction of dark matter that virialized in halos above the minimum $T_{\text{vir}}$. We denote this fraction by $f(z)$ and calculate the timescale for the assembly of cold gas into galaxies from the time derivative $\tilde{f} = \left[\frac{df}{dz}\right]/(dt/dz)$, where $dt/dz = 8.4 \times 10^7 \text{ yr}[(1+z)/10]^{-5/2}$, so that

$$t_{\text{gal}} \equiv f(t) \approx 3 \sqrt{\frac{d\ln f}{d\ln(1+z)}}^{-1} t_{\text{Hub}}.$$  

In the Press-Schechter formalism [31], $f = \text{erfc}[\delta_c(z)/\sqrt{2\sigma(M_{\text{min}})}]$, where $\delta_c = 1.686/D(z)$ is the collapse threshold for an overdensity linearly-extrapolated to $z = 0$, $D(z)$ is the linear growth factor of density perturbations (with $D(0) = 1$), and $\sigma(M_{\text{min}})$ is the root-mean-square amplitude of linearly-extrapolated density perturbations at $z = 0$ on the scale of a sphere from which the minimum galaxy mass is assembled. Refinements to the halo mass function [32] have a negligible effect ($\lesssim 15\%$) on our results for $t_{\text{gal}}$.

**Results.** Figure 1 shows the ratio $(t_{\text{gal}}/t_E)$ as a function of redshift $z$ for the concordance $\Lambda$CDM cosmology [2]. The galaxy growth time is calculated as the fraction of matter that is incorporated in dark matter halos with virial temperatures above $10^9 K$ (dashed line), $10^4 K$ (solid), and $10^2 K$ (dotted). We adopt $\epsilon[(1-\epsilon)] = 32\%(L/L_E)$, corresponding to a BH spin-averaged radiative efficiency of $\epsilon = (6 + 42)/2 = 24\%$ for $L = L_E$ or $\epsilon = 5.7\%$ for the characteristic value [13] of $L/L_E \approx 0.2$. This choice is conservatively smaller by a factor of $\sim 2$ than the lower limit implied by observational data on high-redshift quasars [25]: a corresponding increase in this value by some factor would have lowered the plotted curves by the same factor and strengthened our conclusions. The three vertical lines mark the central value
mergers. For the post-reionization case of is infused into the centers of galaxies through episodic inhibited by the much slower rate at which fresh cold gas early rapid growth. At late times, quasar growth is limited by its cumulative accretion rate, and the Eddington limitation on the accretion time is independent of their cold gas reservoir (aided by the declination merger rate of galaxies). The characteristic time for doubling the mass of galaxies traces the age of the Universe \((\sim 2.3 \times 10^9 \text{ yr}[(1 + z)/4]^{-3/2})\), and keeps increasing relative to \(t_E\) with decreasing redshift. The mode of feedback-regulated growth \([8, 12, 13]\) becomes critical at these low redshifts.

During the early history, when BH growth is not yet regulated by feedback or by the exhaustion of the cold gas reservoir in galactic nuclei, the comoving mass density of accreting BHs \((\rho_{\text{BH}})\) grows at a rate,

\[
\rho_{\text{BH}} = \rho_{\text{BH}}/t_E + \dot{\rho}_{\text{seed}},
\]

where \(\dot{\rho}_{\text{seed}}\) is the formation rate density of BH seeds. If the BH seeds cannot ionize the Universe on their own (i.e., if reionization did not result from the accretion luminosity of stellar-mass BHs), then Fig. 1 shows that stars are required and able \([45]\) to make reionization happen by \(z = 10.9 \pm 1.4\).

At late times, the BH mass density grows exponentially until feedback and the consumption of cold gas (which were not included in Eq. 5) start to regulate its growth \([46]\). By requiring that quasars match the UV production rate per comoving volume of stars only as late as \(z \sim 6–3\) \([1]\), we infer that they fall short of matching it at \(z \sim 11\) by orders of magnitude. The transition redshift between the early domination by stars and the late domination by quasars is sensitive to uncertain parameters (which may also be redshift dependent), such as the unknown feedback strength and the radiative efficiency of these source populations.

The radiative efficiency of stars and BHs is obviously different. The total number of emitted ionizing photons per baryon incorporated into stars ranges \([36]\) between \(4 \times 10^4\) for a present-day (Pop II) mass function and \(\sim 10^5\) for metal-free massive stars (Pop III). Supermassive BHs produce \(\sim 2 \times 10^7\) \(\epsilon\) ionizing photons per accreted baryon \([37]\). The observation that quasars dominate the cosmic UV production rate only at low redshifts \([1]\) implies that their BH formation efficiency out of cold gas in galaxies is lower than that of stars by several orders of magnitude. This is likely caused by angular momentum which distributes the cold galactic gas in a large-scale disk that fragments into stars long before tidal or viscous torques enable a small fraction of this gas to feed the central BH. The inference of a low BH formation efficiency is confirmed by data on the cumulative mass budgets of supermassive BHs and stars in the local Universe \([35]\).

The fundamental limitation presented by Fig. 1 applies also to an early population of stellar-mass black holes, because those would still require extensive accretion in order to reionize the Universe. The comoving luminosity density of any BH population is simply proportional to its cumulative accretion rate, and the Eddington limitation on the accretion time is independent

![Diagram](image-url)

FIG. 1: Redshift evolution of the growth time (in \(t_E\) units) for the fraction of matter that is incorporated in dark matter halos with virial temperatures above \(10^5\) K (dashed line), \(10^6\) K (solid), and \(10^7\) K (dotted). Conservatively, we adopt a value of \(\epsilon/(1 - \epsilon) = 0.32 L/L_E\), which is a factor of \(\sim 2\) smaller than the lower limit implied by observations of high redshift quasars \([25]\). The vertical lines correspond to the central value (solid line) plus or minus one standard deviation (dashed lines) for the redshift of reionization, \(z_{\text{reion}}\), based on the WMAP5 data \([2]\). Prior to reionization, the assembly of cold gas into galaxies occurs at a faster rate than the Eddington-limited growth of BHs.
of the mass distribution of the early BHs. Indeed, the smallness of the unresolved component of the soft X-ray background places severe constraints on the cumulative mass density of an early population of accreting BHs irrespective of its mass distribution [40].

**Observable signatures.** The delay associated with the Eddington-limited growth of black holes in the early universe can be probed through a number of observational methods. First, gravitational wave signals from coalescing BH binaries at high redshifts – which are detectable by LISA [57] and possibly also by Advanced LIGO [38], can be used to probe the mass function of massive BHs as a function of redshift. Second, future X-ray missions such as the International X-ray Observatory [48], and infrared telescopes such as JWST [41], or new ground-based instruments, may detect fainter quasars at higher redshifts than those accessible with present-day telescopes. Third, BHs that escaped from their dwarf galaxies by gravitational-wave recoil at early cosmic times, are potentially detectable in the Milky-Way halo through the compact star clusters that surround them [39]. Finally, extending current simulations of quasar growth [8, 9] to the higher redshifts of reionization (\(z \gtrsim 12\)) would refine predictions for the ionized bubble sizes, which may be probed by future 21cm observatories [37, 41].

**Acknowledgments.** I thank J. Pritchard for helpful comments. This work was supported in part by Harvard University funds.

[1] C. A. Faucher-Giguère, A. Lidz, L. Hernquist, & M. Zaldarriaga, Astrophys. J. **688**, 85 (2008).
[2] E.Komatsu et al. Astrophys. J. Suppl., in press (2008) [arXiv:0803.0547].
[3] A. Loeb, preprint [arXiv:0804.2258] (2008), and references therein.
[4] E. E. Salpeter, Astrophys. J. **140**, 796 (1964).
[5] M. J. Rees, Ann. Rev. Astron. & Astrophys. **22**, 471 (1984).
[6] T. R. Lauer, et al., Astrophys. J., **664**, 226 (2007).
[7] A. Toomre, & J. Toomre, Astrophys. J. **178**, 623 (1972); J. E. Barnes, & L. Hernquist, Astrophys. J. **370**, L65 (1991); T. J. Cox, et al., Mon. Not. R. Astr. Soc. **384**, 386 (2008), and references therein.
[8] T. Di Matteo, V. Springel, & L. Hernquist, Nature **433**, 604 (2005).
[9] Y. Li, et al., Astrophys. J., **665**, 187 (2007).
[10] P. F. Hopkins, L. Hernquist, T. J. Cox, & D. Kereș, Astrophys. J. Suppl. **175**, 356 (2008); J. S. B. Wyithe, & A. Loeb, Mon. Not. R. Astr. Soc., submitted (2008) [arXiv:0810.3155].
[11] K. Gebhardt, et al. Astrophys. J. Lett. **539**, L13 (2000); L. Ferrarese, & D. Merritt, Astrophys. J. Lett. **539**, L9 (2000); S. Tremaine, S., et al. Astrophys. J. **574**, 740 (2002).
[12] J. Silk, & M. J. Rees, Astron. & Astrophys. **331**, L1 (1998).
[13] J. S. B. Wyithe, & A. Loeb, Astrophys. J. **595**, 614 (2003).
[14] A. S. Eddington, The Internal Constitution of Stars, Cambridge: Cambridge University Press (1926).
[15] J. A. Kollmeier, et al. Astrophys. J. **648**, 128 (2006).
[16] S. L. Shapiro, & S. A. Teukolsky, Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects, New York: Wiley (1983), pp. 395-396, 429.
[17] R. Narayan, & E. Quataert, Science **307**, 77 (2005).
[18] C.-L. Jiao, L. Xue, W.-M. Gu, & G.-F. Lu, Astrophys. J., in press (2008) [arXiv:0811.2451].
[19] R. Shaesee, et al., Astrophys. J. Lett. **687**, L25 (2008); R. Shaesee, R. Narayan, R., & J. E. McClintock, Astrophys. J. **676**, 549 (2008).
[20] N. I. Shakura, & R. A. Sunyaev, R. A., Astron. & Astrophys. **24**, 337 (1973).
[21] A. Eigenbrod, et al., Astr. & Astrophys. **490**, 933 (2008); C. W. Morgan, C. S. Kochanek, N. D. Morgan, E. E. & Falco, Astrophys. J. Lett., submitted (2007) [arXiv:0710.0365]; S. Poindexter, N. Morgan, & C. S. Kochanek, Astrophys. J. **673**, 34 (2008).
[22] I. D. Novikov, & K. S. Thorne, Black Holes (Les Astres Occlus), pp. 434-450 (1973).
[23] A. Soltan, Mon. Not. R. Astron. Soc. **200**, 115 (1982).
[24] F. Shankar, D. H. Weinberg, & J. Miralda-Escudé, Astrophys. J., in press (2009) [arXiv:0910.4488].
[25] F. Shankar, M. Croces, J. Miralda-Escude', P. Fosalba, & D. H. Weinberg, Astrophys. J., submitted, (2008) [arXiv:0810.4919].
[26] R. C. Kennicutt, IAU Symp. #237, Cambridge: Cambridge Univ. Press, pp. 311-316 (2007).
[27] X. Fan, New Astron. Rev. **50**, 665 (2006).
[28] R. Barkana, & A. Loeb, Phys. Rep. **349**, 125 (2001).
[29] Z. Haiman, M. J. Rees, & A. Loeb, Astrophys. J. **476**, 458 (1997); Z. Haiman, T. Abel, & M. J. Rees, Astrophys. J. **534**, 11 (2000); M. Dijkstra, Z. Haiman, A. Mesinger, & J. S. B. Wyithe, Mon. Not. R. Astr. Soc., in press (2008) [arXiv:0810.0014], and references therein.
[30] J. S. B. Wyithe, & A. Loeb, Nature **441**, 322 (2006), and references therein.
[31] W. H. Press, & P. Schechter, Astrophys. J. **187**, 425 (1974).
[32] R. K. Sheth, & G. Tormen, Mon. Not. R. Astr. Soc. **329**, 61 (2002); A. Jenkins et al., Mon. Not. R. Astr. Soc. **321**, 372 (2001); M. S. Warren, K. Abazajian, D. E. Holz, & L. Teodoro, L., Astrophys. J. **646**, 881 (2006); J. L. Tinker et al., Astrophys. J., submitted (2008) [arXiv:0803.2706].
[33] T. Tanaka, & Z. Haiman, Astrophys. J., submitted (2008) [arXiv:0807.4702].
[34] R. M. O’Leary, & A. Loeb, MNRAS, submitted (2008) [arXiv:0809.4262].
[35] J. Magorrian, et al., Astr. J. **115**, 2285 (1998); S. Tremaine, et al., Astrophys. J. **574**, 740 (2002); N. Haring, & H.-W. Rix, Astrophys. J. **604**, L89 (2004).
[36] V. Bromm, R. P. Kudritzki, & A. Loeb, Astrophys. J. **552**, 464 (2001).
[37] J. S. B. Wyithe, A. Loeb, & C. Carilli, Astrophys. J. **628**, 575 (2005); R. C. Telfer, W. Zheng, G. A. Kriss, & A. F. Davidsen, Astrophys. J. **565**, 773 (2002); R. L. White, R. H. Becker, X. Fan, X., & M. A. Strauss, Astr. J. **126**, 1 (2003).
[38] J. S. B. Wyithe, & A. Loeb, Astrophys. J. **612**, 597
(2004); *ibid* **590**, 691 (2003).

[39] R. M. O’Leary, & A. Loeb, Mon. Not. R. Astron. Soc., submitted (2008) [arXiv:0809.4262].

[40] M. Dijkstra, Z. Haiman, & A. Loeb, Astrophys. J. **613**, 646 (2004).

[41] A. Loeb, preprint (2008) [arXiv:0804.2258].

[42] The luminosity of star forming galaxies is typically well below the Eddington limit for their total mass.

[43] It is possible that the most massive BHs at $z \sim 6$ have formed with a lower radiative efficiency than the bulk of the BH population, allowing them to grow more rapidly at earlier cosmic times. This might be an important selection bias, since the quasars would not have been detected by existing surveys [27] if they were fainter.

[44] UV photons are produced by massive stars with lifetimes well below the age of the Universe at the redshifts of interest (hundreds of millions of years). Even when a massive star radiates near the Eddington limit, its lifetime is only a few million years because its radiative efficiency is much lower than that of a BH (see Eq. 2). Despite their lower efficiency, stars remain competitive with BHs as UV sources because they generically consume a much bigger fraction of the baryons in galaxies [11].

[45] Aside from the Eddington limit, the growth rate of the early BH population could have been inhibited by gravitational wave recoil of merger remnants out of the shallow potential wells of the first dwarf galaxies [33, 34].

[46] This regulation would lead to a saturation value of $\rho_{BH}$ that is proportional to the comoving density of stars $\rho_\star$ at very late times, as indicated by observations [35].

[47] http://lisa.nasa.gov/

[48] http://ixo.gsfc.nasa.gov/

[49] http://www.stsci.edu/jwst/