# Galaxy Assembly and SMBH/AGN-growth from Cosmic Dawn to the End of Reionization


## Paul Scowen

**Research Professor**

**School of Earth & Space Exporation**
**Arizona State University**
**PO Box 876004, Tempe, AZ 85287-6004**
**(480) 965-0938**
***paul.scowen@asu.edu***

**Rolf A. Jansen (Arizona State University, *rolf.jansen@asu.edu*)**

**Rogier Windhorst (Arizona State University, *rogier.windhorst@asu.edu*)**

**James Rhoads (Arizona State University, *james.rhoads@asu.edu*)**

**Sangeeta Malhotra (Arizona State University, *sangeeta.malhotra@asu.edu*)**

**Daniel Stern (NASA/Jet Propulsion Laboratory, *daniel.k.stern@jpl.nasa.gov*)**

**Robert O'Connell (University of Virginia, *rwo@viginia.edu*)**

**Matthew Beasley (University of Colorado – Boulder, *beasley@casa.colorado.edu*)**

**for the *HORUS* & *SFC* science concept teams**


## Science RFI Response to NASA Cosmic Origins Program

**Abstract:** In order to address the key Cosmic Origins science question ***"How did galaxies evolve from the very first systems to the types we observe nearby?"***, we propose to the community a *systematic* and *comprehensive* UV–near-IR cosmological broad- and medium-band imaging and grism survey that covers a wide area on the sky in muliple epochs. Specifically we advocate a tiered survey that covers $\sim$10 deg$^2$ in two epochs to $m_{AB} \sim 28$ mag, $\sim$3 deg$^2$ in seven epochs to $m_{AB} \sim 29$ mag, and $\sim$1 deg$^2$ in 20 epochs to $m_{AB} \sim 30$ mag, each at 10$\sigma$ point source sensitivity. Such a survey would provide spectrophotometric redshifts accurate to $\sigma_z/(1+z) \lesssim 0.02$ and faint source variability measurements for $\gtrsim 5\times10^6$ galaxies and QSOs, and would be an essential complement to *JWST* surveys ($\lesssim 0.1$ deg$^2$ to $m_{AB} \lesssim 31$ mag at $\lambda >$1100 nm). This rich data set would allow: (1) study of faint Ly$\alpha$-emitting and Lyman-break galaxies at $5.5 \lesssim z \lesssim 8$ to understand how galaxies formed from primordial density perturbations and to trace the metal-enrichment of the intergalactic medium (IGM); (2) measuring the evolution of the faint end of the galaxy luminosity function (LF) from $z \sim 8$ to $z \sim 0$ by mapping the ramp-up of Pop II star formation, (dwarf) galaxy formation and assembly, and hence, the objects that likely completed the Hydrogen reionization by $z \simeq 6$; (3) direct study of the $\lambda <$ 91.2 nm escape fractions of galaxies and weak AGN from $z \sim 4.0$–2.5, during the epoch of Helium reionization; (4) measuring the mass- and environment-dependent galaxy assembly process for $\gtrsim 5\times10^6$ galaxies from $z \simeq 5$ to $z \simeq 0$; (5) tracing the strongly epoch-dependent galaxy merger rate and constraining how Dark Energy affected galaxy assembly and the growth of super-massive black holes (SMBHs); (6) the study of $\gtrsim 10^5$ weak AGN, including faint variable objects (feeding SMBHs in the faint end of the AGN LF), over 10 deg$^2$ to measure how SMBH growth kept pace with galaxy assembly and spheroid growth, and how this process was shaped by various feedback processes over cosmic time. The proposed study is not feasible with current instrumentation but argues for a wide-field ($\gtrsim$250 arcmin$^2$), high-resolution ($\lesssim 0\farcs1$), UV–near-IR imaging facility on a 2.4–4 m class space-based observatory.

**O**ver the past decade, our knowledge about the universe at high redshifts has gradually extended to $z \simeq 7$ with dozens of quasars discovered in the *SDSS*, *UKIDDS*, and *CFHT-LS* surveys at $z \gtrsim 6$ and similar numbers of Ly$\alpha$ emitters. Of particular note are the discoveries of the first "complete" Gunn-Peterson troughs in the spectra of $z>6$ quasars and the WMAP year-7 polarization measurement, which gives a 2$\sigma$ upper limit to the redshift range of the Pop III star reionizing population of $z \simeq 8$–14. The reionization of the universe likely has left its signature on the history of galaxy formation and evolution. It is predicted to cause a drop in the cosmic star formation rate (SFR), and is therefore accompanied by a dramatic fall in the number counts of objects at $z \geq 6$.

Since the UV shortward of $\lambda_0$=121.6 nm is strongly absorbed by intervening H I, high redshift objects can be selected using the so-called *drop-out* technique. This technique requires filters that bracket Ly$\alpha$ in the relevant redshift range. Recent *i*-, *z*-, *Y*-, and *J*-band drop-out studies with *HST* found significant numbers of $z \simeq 7$ candidates, although with non-negligible contamination by low redshift elliptical galaxies and Galactic L- and T-dwarf stars. As evidence mounts that the Hydrogen reionization was largely complete by $z \simeq 6$, studies of the $z = 6$–8 interval — "Cosmic Dawn" —, will be of great cosmological importance.

For how galaxies formed from perturbations in the primordial density field, reflected in the Cosmic Microwave Background (CMB), remains a major problem. While numerical simulations can predict the formation of dark matter halos and their clustering, the formation of stars that render these halos visible is a complex process and hard to predict *a priori*. Thus, there is a great need to

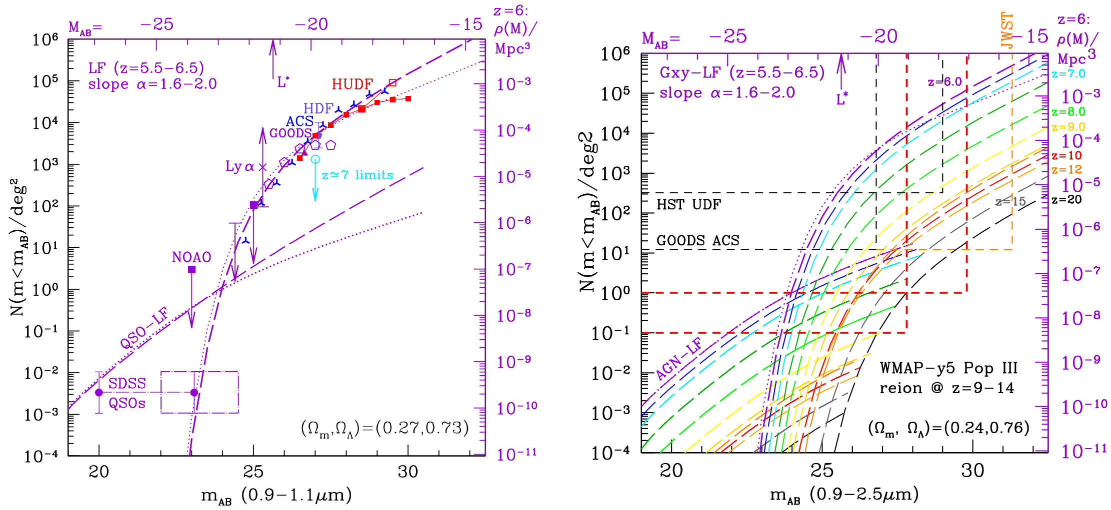


Fig. 1 — (*a*) The integral luminosity function at $z \simeq 6$ from various samples. Left and bottom axes give the observed surface densities and fluxes, top and right axes absolute magnitudes and space densities. The surface density of $z \simeq 6$ QSOs to AB=20 mag from *SDSS* and its extrapolation to fainter fluxes using the measured faint-end slope of the QSO LF ($|\alpha| \simeq 1.6$) are also shown. The faint-end slope of the galaxy LF is significantly steeper: $|\alpha| \simeq 1.8$–2.0 (long-dashed curve). (*b*) The predicted surface density of $z \simeq 7$–8 objects, based on the few candidates from the *HST*/NICMOS and WFC3 surveys (*light blue upper limit in panel* a) and constrained by the total optical depth $\tau$ from WMAP, may be an order of magnitude lower than that at $z \simeq 6$, necessitating the very large survey areas proposed here (*red dashed limits*).

study galaxies observationally, at all redshifts. This is especially true at $z \gtrsim 6$, where two major changes took place: (*1*) metal enrichment of the intergalactic medium (IGM), which must have occurred at $z \gtrsim 6$ given the observations of IGM metals even at $z$=5.7, and (*2*) reionization of hydrogen in the IGM. Since metallicity and ionization change the nature of star formation by changing the available cooling mechanisms, it is *crucial* to push back galaxy samples to $z > 6$.

Surveys for galaxies at $z \gtrsim 7$ are very difficult for many reasons, however. The galaxies are fainter, both because of cosmological dimming and also because of smaller characteristic luminosities and sizes, resulting in low object surface densities (e.g., Fig. 1). It is also important to realize that high redshift galaxy formation is *biased*, resulting in strong spatial variations in number density. For these reasons one would need to survey a large area (at least several deg$^2$). These searches need to be performed at $\lambda \gtrsim 975$ nm, near and beyond the cut-off of Si CCDs. In the near-IR, there is a tremendous advantage of going to space, with its >100–1000 times darker sky background.

One class of primordial galaxies is easily identified in narrow- or medium-band surveys from their strong, narrow Ly$\alpha$ emission and their diminished flux blueward of this emission. Indeed, Ly$\alpha$-emitter surveys have proven to be *the* most successful technique to find galaxies at the earliest cosmic epochs. While the Gunn-Peterson troughs are produced by neutral fractions of only $10^{-4}$ or $10^{-2}$ (for a homogenous or a clumpy IGM, respectively), the change in number density of Ly$\alpha$-emitters as a function of redshift traces neutral fractions of the IGM of $\gtrsim$30–80%. A quantitative study based on this principle requires statistical samples of Ly$\alpha$ galaxies in each redshift bin. Ground-based surveys are and will remain severely limited in the volume they can sample due to the necessity to use very narrow bandpass filters (~0.1%) to observe between the strong atmospheric OH lines, which makes them vulnerable to cosmic variance.

The Hubble Ultra-Deep Field (HUDF; Beckwith et al. 2006), our deepest view yet of the distant universe, was collected over 4 epochs that were each ∼1 month apart. Since the data of each seperate epoch still reaches to ∼28 mag, this offered the unique opportunity to study the variability of faint objects on time scales of months, corresponding to 4–5 weeks in the rest-frame. Variability on such time scales betrays the presence of a feeding SMBH within a galaxies' active nucleus (AGN). The redshift distribution of galaxies in a particular early-merger stage and variable objects in the HUDF appear to be similar, which may be a clue to the mystery of how the growth of spheriods and SMBHs has kept pace with the process of galaxy assembly and resulted in the tight Magorrian-relation observed locally. The present statistics are inadequate, however, and the available redshift estimates imprecise. A deep, multi-epoch survey over $\gtrsim$1 deg$^2$ would allow studying variability of faint objects over a 1000× larger area on the sky to similar depth, providing vastly superior statistics.

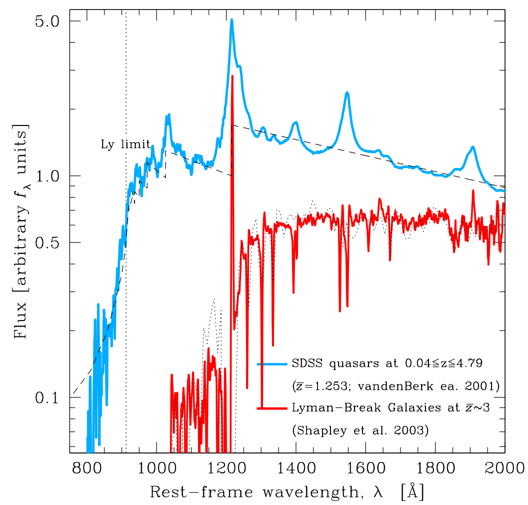


Fig. 2 — Whether the numerous dwarf galaxies or the much rarer AGN finished reionization at $z \simeq 6$ depends critically on the amplitude and faint-end slope of the LF for each population. The steep LF of dwarf galaxies at $z \sim 6$ could provide enough ionizing photons to complete reionization if $f_{esc} \gtrsim$ 10%. Observations in UV–blue filters in the proposed survey will yield escape fractions for $\gtrsim$5×10$^6$ objects.

While *HST* has meant a quantum leap forward, efficient access of the high-$z$ universe is still severely limited by the product $\Omega \times A \times \Delta\lambda$, of small field of view (FoV), limited aperture, and limited wavelength range over which it provides high throughput. A comprehensive study of the galaxy populations from the height of the reionization epoch to the epoch where the present-day Hubble sequence was established, would require a space-based imaging facility that provides:

***(1)*** efficient wide-field coverage ($\gtrsim$250 arcmin$^2$), sufficient to efficiently map areas large enough to average out cosmic variance and find $z \gtrsim 7$ objects with surface densities $\gtrsim$0.1/deg$^2$;

***(2)*** high angular resolution, sufficient to spatially resolve ∼1 kpc sized objects at $0.5 \lesssim z \lesssim 8$ at restframe wavelengths $\lambda_0 >$121.6 nm;

***(3)*** sufficient sensitivity to sample both the bright and faint ends of the LFs of galaxies, QSOs, Ly$\alpha$-emitters and Lyman-break objects from $z \simeq 8$ to $z \simeq 1$, and to $z \simeq 0$ for the Balmer or 400 nm breaks; *and*

***(4)*** a sufficiently rich complement of near-UV–near-IR broad- and medium-band filters to provide photometric redshift estimates accurate to $\sigma_z/(1+z) \lesssim 0.02$ and to allow efficient detection of Ly$\alpha$-emitters from $z \simeq 8$ to $z \simeq 5.5$ .

We therefore propose to the community a near-UV–near-IR cosmological broad- and medium-band imaging and grism survey that covers a wide area on the sky in muliple epochs. Specifically we advocate a tiered survey of ∼10 deg$^2$ in two epochs to $m_{AB}$ = 28 mag, ∼3 deg$^2$ in seven epochs to $m_{AB}$ = 29 mag, and ∼1 deg$^2$ in 20 epochs to $m_{AB}$ = 30 mag, each at 10$\sigma$ point source sensitivity. The use of complementary deep, medium-deep, and wide surveys is a proven strategy to maximize the scientific return for the investment in telescope time. Such surveys would provide spectrophotometric redshifts accurate to $\sigma_z/(1+z) \lesssim 0.02$, faint source variability for $\gtrsim 5\times10^6$ galaxies and QSOs, and a probe of the universe at Cosmic Dawn when less than half of the hydrogen had been ionized. It would constitute an essential complement to deeper *JWST* surveys $m_{AB} \lesssim 31$ mag at $\lambda > 1100$ nm and $z \gtrsim 8$) over far smaller areas ($\lesssim 0.1$ deg$^2$).

In the following, we summarize the science goals for each of the themes of this survey.

**Key scientific themes that have arisen from recent advances**

***Evolution of the Faint-end Slope of the Dwarf Galaxy Luminosity Function*** The faint-end slope of the galaxy LF is systematically steepening at higher redshifts, reaching a slope $|\alpha|$=1.8–2.0 at $z \sim 6$. This implies that dwarf galaxies collectively could have produced a sufficient number of ionizing photons to complete the reionization of Hydrogen in the universe by $z \sim 6$. This critically depends on the escape fraction, $f_{esc}$, of far-UV photons from faint dwarf galaxies. The proposed survey, in particular the UV–blue broad-band filters, could answer this question for statistically meaningfull samples per redshift bin. It furthermore depends on the evolution of the amplitude of the dwarf galaxy LF and whether or not there could be a significant scatter in the faint-end slope due to clustering. The surface density of $z > 7$ objects appears to be an order of magnitude lower than that at $z \sim 6$, but these wide-area surveys would unambiguously answer these questions.

***Tracing the Reionization History using Ly$\alpha$-Emitters*** Observations so far have failed to settle the issue of whether the amplitude of the Ly$\alpha$-emitter LF changes between $z = 5.7$ and $z = 6.5$, or as extrapolated from objects and candidates at $z \gtrsim 7$. The proposed medium-band surveys will derive their LF as a function of redshift at $z \gtrsim 5.5$ over a wide area for large statistical samples and definitively address how the reionization of the IGM progressed over time. Furthermore, the data will allow measuring the ages and clustering properties of Ly$\alpha$-emitters, and, via the faint-end slope of their LF, their contribution to the budget of ionizing photons. The latter is a complementary probe of cosmic reionization compared to the counting experiment (LF amplitude).

***Light Profiles of Dwarf Galaxies Around Reionization*** The average radial surface brightness profile derived from stacked, intrinsically similar, $z \simeq 6$, $z \simeq 5$, and $z \simeq 4$ objects extracted from the *HST*/ACS HUDF show a deviation from a Sersíc profile at progressive larger radii. If interpreted as a virial radius, in a hierarchical growth scenario, this would imply dynamical ages for these dwarf galaxies of a 0.1–0.2 Gyr at $z \simeq$ 6–4. These 'dynamical' limits to their ages are comparable to age estimates based on their SEDs, suggesting that the starburst that *finished* the H reionization at $z \simeq 6$ may have started by a global onset of Pop II star formation at $z \simeq$ 6.5–7, or $\lesssim 200$ Myr before $z \simeq 6$. The proposed surveys will yield light profiles, color gradients, and dynamical states of $\gtrsim 10^5$ dwarf galaxies at $0.5 \lesssim z \lesssim 7$, and provide constraints to their ages from their SEDs and, for a subset, also from systematic profile deviations.

***Lyman-continuum Escape Fraction of Dwarf Galaxies and Weak AGN*** At $z \simeq 6$, the Lyman-continuum escape fraction is likely somewhat larger than the 10–15% measured for Lyman-break galaxies at $z \simeq$ 3–4, reflecting the lower metallicity at larger redshifts. If indeed dwarf galaxies, and not QSOs, dominated the late stages of reionization, then these objects cannot have started shining pervasively much before $z \simeq$ 7–8, or no neutral H I would have been detected in front of $z \simeq 6$ *SDSS* quasars. Hence, one would expect to find a down-turn in their LF amplitude at $z \gtrsim 6.5$ — or a rapid onset of the cosmic SFR from $z \simeq 8$ to $z \simeq 6$, which may be identified with the onset of dwarf galaxy formation. The proposed surveys will provide a unique glimpse into this era of 'Cosmic Dawn', where the first global IMF of Pop II stars in dwarf galaxies started forming.

***The Process of Hierarchical Galaxy Assembly*** The process of galaxy assembly may be directly traced as a function of mass and cosmic environment in the redshift range $0.5 \lesssim z \lesssim 5$. The *HST* Deep Fields have outlined how galaxies formed over cosmic time, by measuring the distribution over structure and type as a function of redshift. Sub-galactic units appear to have rapidly merged from $z \simeq 6$–8 to grow bigger units to $z \simeq 1$. Galaxies of all types formed over a wide range of cosmic time, but with a notable transition around $z \sim 1.0$. Merger products started to settle as galaxies with familiar morphologies, and evolved mostly passively since then. The fine details of this process still elude the *HST* surveys, because of inadequate spatial sampling and/or depth, and because its FoV is too small to provide sufficient statistics. The proposed imaging through multiple near-UV–near-IR filters and grism(s) would yield robust spectrophotometric redshift estimates for $\gtrsim 5\times10^6$ galaxies with $m_{AB} \lesssim 28$–30 mag, and allow an analysis of their stellar populations (through population synthesis modeling) and their structure on spatial scales $\lesssim$ few kpc.

***The Epoch-dependent Merger Rate of Galaxies*** With robust photometric redshift estimates, it has become feasible to meaningfully trace the pair fraction and galaxy major merger rate to very faint limits ($m_{AB} \gtrsim 27$ mag). From *HST*/ACS flux limits and panchromatic SED fitting, the currently available surveys have shown a mass completeness limit for $z \lesssim 2$–4 for $M \gtrsim 10^{10.0}\,M_\odot$ for primary galaxies in a pair and $M \gtrsim 10^{9.4}\,M_\odot$ for secondary galaxies. The proposed surveys would allow mapping the *entire* epoch-dependent merger history to at least 3 mag fainter. This would yield the galaxy merger density as a function of total mass, mass ratio, redshift, and local overdensity and do so for $\gtrsim 10^6$ galaxies at $m_{AB} \lesssim 28$–30 mag over a much wider range of masses ($10^{9.8}\,M_\odot \lesssim M \lesssim 10^{11.5}\,M_\odot$) and for redshifts $0 \lesssim z \lesssim 7$.

***The Growth of Super-Massive Black Holes*** Through a multi-epoch variability study, the proposed surveys will be able to measure the weak AGN fraction in $\gtrsim 10^5$ field galaxies to $m_{AB} \lesssim 28$–30 mag at $z \lesssim 8$ directly, and so robustly constrain how exactly growth of spheroids and SMBHs kept pace with the process of galaxy assembly. The panchromatic imagery and robust spectrophotometric redshifts will allow decomposition of the AGN light from that of the underlying galaxy. This science theme also relies on a stable PSF and proper PSF sampling.

**Broad Design Considerations Driven by this Science**

*Resolution* — In order to spatially resolve kpc-sized objects at $0.5 \lesssim z \lesssim 8$ at rest-frame wavelengths $\lambda_0 > 121.6$ nm, a resolution $\lesssim 0\farcs04$ would be required.

*Wavelength agility* — pan-chromatic wavelength coverage from near-UV through near-IR for a comprehensive understanding of the star-formation and assembly histories of galaxies, and to access Ly$\alpha$ emission redshifted to $z \sim 8$.

*Wide-field focal plane arrays* — these are presently not at sufficiently high TRL; investment is needed to improve yields, provide cheaper devices and high-throughput assembly and testing to enable economies of scale. (Such an investment would not just benefit the science proposed here)

*Coatings* — an investment in improving the relatively poor broad-band performance of optical coatings of telescope mirrors in the UV, with typical reflectances below 85% (Al+$MgF_2$) directly results in a large increase in throughput for a given telescope aperture, or more affordable missions for a given sensitivity requirement.

*Dichroics* — most photons collected by telescopes are rejected by bandpass filters. Dichroic(s) potentially double (or even triple) the observing efficiency of astronomical observatories (e.g., *Spitzer*/IRAC) and allow tuning downstream optics and detectors for more optimal performance, avoiding compromises inherent in forcing performance over more than an octave in frequency.

The proposed science program does not stand alone, but must build on a strong understanding of the physics of the star formation process in various environments, theoretical insights in cosmological models of reionization and structure growth, as well as synergy with both higher-resolution near-IR AO observations with next-generation giant-aperture telescopes, and deeper observations in the near- and mid-IR with *JWST* over small fields of view.

Combination of a large collecting area, very wide field of view, high angular resolution, wavelength agility and/or multiplexing advantage would allow orders of magnitude more efficient UV–optical observations of star formation, galaxy assembly, and SMBH-growth processes and, moreover, open up a new domain in discovery space near and far. Injection into L2 (or an Earth Drift-Away orbit) provides dynamical and thermal stability, and doubles efficiency over LEO orbits and, hence, lowers cost per hour of observation (all other variables being equal). Large focal plane array (dozens to hundreds of individual CCD or CMOS detectors) and dichroic camera (simultaneous observation in two or more channels of the same field of view) technology is better matched to the collimated beams provided by optical telescope assemblies and less wasteful in terms of collected photons, maximizing science output and especially benefitting survey science with a lasting legacy beyond the nominal duration of a mission. Survey science allows discovery of very rare objects amongs billions and billions, the positions an properties of which may not be knowable a priori.

**Four central questions to be Addressed**

***(1)*** How did reionization progress during the era of 'Cosmic Dawn'? Was it an extended, a rather abrupt, or even a multiple event?

***(2)*** How did the faint end of the galaxy luminosity function evolve from the onset of Pop II star formation till the end of the reionization epoch?

***(3)*** How exactly did AGN and SMBH growth keep pace with the process of galaxy assembly? How did AGN growth decline with the galaxy merger rate and the cosmic SFR?

***(4)*** Was there indeed an epoch of maximum merging and AGN activity around $z \simeq 1$–2 for the more massive galaxies, before the effects from the increasingly dominant Dark Energy kicked in? How does this peak epoch depend on galaxy total mass or bulge mass, and (how) does this support the galaxy downsizing picture?