First Light and Reionization :
open questions in the post-JWST era

Massimo Stiavelli
Phone : 410-338-4835
Space Telescope Science Institute
mstiavel@stsci.edu

and

John Mather (NASA/GSFC, chair)
Mark Clampin (NASA/GSFC)
Rene Doyon (U. of Montreal)
Kathy Flanagan (STScI)
Marijn Franx (Leiden U.)
Jonathan Gardner (NASA/GSFC)
Matthew Greenhouse (NASA/GSFC)
Heidi Hammel (SSI)
John Hutchings (Herzberg I. of A.)
Peter Jakobsen (ESA)
Simon Lilly (ETH-Zurich)
Mark McCaughrean (U. of Exeter)
Matt Mountain (STScI)
George Rieke (U. of Arizona)
George Sonneborn (NASA/GSFC)
Rogier Windhorst (Arizona State U.)
Gillian Wright (UK ATC)

(the JWST Science Working Group)

1. Introduction

The aim of this paper is to outline the expected JWST performance in addressing first light and reionization science questions that are found to be of interest today. These are some of the most challenging and interesting questions in modern astronomy, and are key drivers for the design of the JWST. Nevertheless, because these early epochs are difficult to observe, even JWST is unlikely to provide complete answers.

Following recombination, the Universe undergoes a phase of expansion and cooling while perturbations initially seeded by dark matter are slowly growing. Small halos are the most common but they have the smallest virial temperatures and require the Universe to be colder than their virial temperature in order to be able to cool radiatively. Thus, it will be the smallest halos able to cool by molecular hydrogen that will form the first structures (Tegmark et al. 1997). In a standard CDM model, these halos form Population III stars at a redshift of 50-60 (e.g. Gao et al. 2005, Naoz et al. 2006, Trenti & Stiavelli 2007). These sources represent the first discrete sources of light. Formation of Population III stars can continue until chemical enrichment makes further formation of these objects impossible. Thus, it is possible that some Population III stars may continue to form down to relatively low redshifts ($z<6$) in isolated pockets of primordial gas. Population III stars are likely massive and with high effective temperature so that they are effective at ionizing hydrogen. Thus reionization is started by Population III stars. It is likely that radiative feedback from other Population III stars self-regulates their formation so that these objects never become the dominant reionization source, in agreement with the Compton optical depth measured by WMAP.

The early formation stages after the onset of cooling by molecular hydrogen are relatively well understood but the later stages of the stellar collapse defining the initial mass function are not theoretically as robust. The star formation history of Population III depends on how effective the Lyman-Werner (LW) radiative feedback really is (e.g. Ricotti et al. 2001, 2002, Machacek et al. 2003, Trenti & Stiavelli 2008, O'Shea & Norman 2008). Moreover, most theoretical results on Population III formation assume a standard CDM. Results would be radically different if dark matter had a different power spectrum. Observational results on these objects would therefore place constraints on the dark matter power spectrum at scales not otherwise easily testable. A first question is then: **When and how did the first stars form?**

Population III stars may leave black hole remnants with mass around 100 $M_\odot$ (Heger & Woosley 2002) and these black holes may seed the formation of the first AGNs (Madau & Rees 2000). A difficulty in this scenario is the effect of black hole mergers which tends to eject black holes from halo centers even in binary mergers, due to gravitational recoil (Volonteri 2007). For this reason, alternative models for the formation of more massive seed black holes have been proposed (Begelman et al. 2006). A second question is then: **When and how did the first Active Galactic Nuclei form?**

Following the formation of the first stars, the first star clusters and the first galaxies will form. It is likely that these objects will not be made of Population III stars but of more evolved metal-enriched stars. Various lines of argument based on numerical simulation (e.g. Wise & Abel 2008) or on the metallicity where cooling by metals becomes important (e.g. Bromm et al. 2001a) suggest a metallicity around 10^{-3} – 10^{-4} solar for these objects. The relevant question in this context is: **What are the first galaxies?**

Evidence from WMAP suggests an epoch of reionization of z=11+/- 1.4 (Komatsu et al. 2009). However, the analysis of the spectra of SDSS QSOs at z=6 would suggest that reionization is completed only at z=6 (Fan et al. 2006). Evidence from Lyman α emitters indicates that the neutral fraction is relatively low (~10-30%) already at z=6.5 (Malhotra & Rhoads 2006). Taken together these indications suggest that reionization has a relatively complex history so that we are led to ask **When and how did reionization occur?**

Following the partial ionization by Population III stars (e.g. Ricotti et al. 2002, Venkatesan et al. 2003), the ionization of hydrogen is completed by some combination of AGNs, and metal-enriched stars. Preliminary evidence from redshift 6 galaxies has failed to identify a galaxy population sufficiently bright to keep the Universe ionized at z=6 assuming relatively normal properties for these objects (Bunker et al. 2004, Yan & Windhorst 2004, Stiavelli et al. 2004, Bouwens et al. 2007). However, the contribution from AGNs appears to be modest and if it wasn't, their harder continuum would reionize also helium which is instead reionized at z~3 (Jakobsen et al. 1994, Davidsen et al. 1996). Thus, presently we still don't know **What sources caused reionization?**

Summarizing, we can identify five high level questions in this area, namely in order of increasing redshift:

1. When and how did reionization occur?
2. What sources caused reionization?
3. What are the first galaxies?
4. When and how did the active galactic nuclei form?
5. When and how did the first stars form?

The following sections will discuss in more detail each question and the expected contribution of JWST.

2. When and how did reionization occur?

Major progress in this area should take place even before JWST. Observations of the 21cm line at high redshifts by LOFAR and MWA should be able to begin constraining the history of reionization by tracking the density of neutral hydrogen as a function of redshift (Shaver et al. 1999, Furlanetto & Loeb 2002).

The study of Lyman α emitters at z>6 is likely to continue in the near future and might place even stronger constraints on the history of reionization as the existing data at z=6.5 are marginally compatible with reionization being completed at z=6 (e.g. Shull & Venkatesan 2008). The discovery of bright Lyman α sources at even higher redshift would not be easily compatible with the idea that reionization is completed at z=6 and would indicate either evolution of the Lyman α forest or strong evolution in the physical or clustering properties of the Lyman α sources. Moreover, several projects to discover QSOs at z>7 are ongoing and might provide further probes for the state of the IGM at high redshift. Finally, Planck and possibly new balloon experiments should be able to further refine the measurement of the Compton optical depth providing an integrated but very powerful constraint on reionization.

On this basis it is well possible that this question will be answered before JWST flies. If it isn't, JWST has the capabilities to address many aspects of this problem. In the study of Lyman α emitters JWST will be able to detect Balmer lines in the brightest ones. The evolution of the Hα luminosity function would avoid the complexity of Lyman α radiative transfer and provide an indication on the intrinsic evolution of their properties so as to rule out or confirm strong evolution. The sensitivity of JWST will allow us to probe the neighborhood of bright Lyman α emitters to determine the role of clustering in setting the size of ionized bubbles. This could be completed by studies of Lyman α emitters in ionized bubbles identified at 21 cm.

Various groups have attempted to study reionization through the Lyman α signature imprinted in the cosmic near-IR background (Baltz et al. 1998, Shaver et al. 1999). This is a very challenging measurement with a wide range of model predictions. If reionization is a slow process (as suggested by existing data), the Lyman α signature would be washed out in redshift and much more difficult to identify. Whether this background has already been detected is not universally agreed upon (Kashlinsky et al. 2007, Thompson et al. 2007). In the near future WFC3-IR and then JWST are in the position to definitely confirm these results or, alternatively, to place strong limits on this signal.

Finally, many proposed experiments seek to identify QSOs at z>7. JWST would be able to obtain high S/N spectra of these objects to study the properties of the IGM at high redshift and further constrain the history of reionization.

On the basis of these considerations we think that the question of **When and how did reionization occur** will be largely solved in the post JWST-era.

3. What sources caused reionization?

Thanks to the UDF and GOODS we have identified more than 500 candidate galaxies at z=6 with a luminosity function extending almost three magnitudes below $L_\star$. Spectroscopic identification is missing for the majority of these objects but the sample appears not to suffer from major contamination for AB<27 and it is likely that the contamination by interlopers or brown dwarfs at fainter magnitudes will be even lower. The ionizing output of these objects needs to be estimated assuming an SED and an escape fraction. When values reasonable for a local population are adopted the objects at z=6 appear inadequate to reionize or to keep ionized the Universe. In order to do so one would have to assume a combination of higher escape fraction, a luminosity function extending to lower luminosity than observed in the UDF and with a steep slope, or a higher ionizing efficiency due to, e.g., very low metallicity. JWST will be able to clarify the nature of these objects by obtaining spectra for the brightest ones. The metallicity could be measured from line ratios and when coupled to the observed continuum would enable one to estimate the ionizing output of a galaxy. Modeling of the Balmer lines and Lyman α would allow one to estimate the fraction of ionizing output producing recombination lines. The difference between these values would yield the escaping ionizing flux.

Very little is known at z=7 or higher. NICMOS investigations have been limited by the small field of view of this instrument. Indications so far are that galaxies at z=7 are rarer by a factor 3 or so than those at z=6 but this result still suffers from small number statistics and is possibly affected by cosmic variance (Yan et al. 2005, Bouwens et al. 2008, Oesch et al. 2009). The installation of WFC3-IR on HST will provide a major improvement to our capabilities and it is likely that a sample of at least a few tens of z=7 galaxies will be identified. However, JWST is probably required to push the study to higher redshift.

It is likely that by identifying faint sources down to AB~31 and studying the spectral properties of galaxies down to AB~28-29, JWST will identify the sources responsible for completing reionization so that even this will be a largely solved problem in the post-JWST era.

4. What are the first galaxies?

The definition of "first galaxy" is ambiguous and a number of techniques will be needed to meaningfully identify these objects. Below we describe a few possible methods (for an expanded discussion see the SWG white paper on this topic at http://www.stsci.edu/jwst/science/whitepapers/first_light_study_V.pdf):

- LF evolution. Models predict that the Luminosity Function (LF) should evolve significantly for the first galaxies. The break in their LF at luminosities $L=L_\star$ will be much fainter than the present value, and, perhaps the slope of the mass function will become steeper at lower masses and fainter luminosities (Bouwens et al. 2007, Yan & Windhorst 2004, Ryan et al. 2007). Equally importantly the value of the density of galaxies $\phi_\star$ is expected to decrease significantly at higher redshifts. We should aim at detecting both a change in the LF slope and characteristic luminosity $L_\star$ and a change in the number density of objects.
- Metallicity. The first galaxies should have much lower metallicity than other objects. However, their metallicity may be non-zero because of self-enrichment or pre-enrichment by Population III stars.
- Absence of an older stellar population. The First galaxies should not have older stellar populations. They will have SED ages no older than their first burst of star-formation.

To further these goals, JWST has been designed to study the luminosity function to the same relative depth as that measured in the UDF at z=6 (i.e. 3 magnitudes below $M_\star$) up to z=20. With this sensitivity JWST will be able to distinguish between luminosity and number evolution and enable a detailed comparison of observations and models. As an example, based on dark halo statistics, a JWST deep field reaching down to AB~31 would identify several tens of galaxies at z>10 (Trenti & Stiavelli 2008).

The sensitivity of NIRSpec (3×10^{-19} erg s^{-1} cm^{-2} in 10^5 s) is such that JWST will be able to measure metallicity of galaxies, using the OIII] 1665 line down to 10^{-3} solar for galaxies 1.5 magnitudes below $M_\star$. This metallicity appears to be close to the metallicity of gas in first generation galaxies and would enable us to establish whether the galaxies we observe are evolved systems. Finally MIRI could be used to identify older stellar population on the basis of the SED of the objects.

It is certain that JWST will be able to push the study of galaxies to much higher redshift than presently possible and it is likely that the nature of the first galaxies will be effectively explored by JWST.

5. When and how did the first active galactic nuclei form?

Some Population III stars may leave black hole remnants with mass of ~100 $M_\odot$. Ionizing radiation from a Population III star is sufficient to expel all gas from the predicted small halo even without a supernova explosion so we should expect a delay after the end of life of the star before the black hole can begin accreting. The luminosity of the accreting black hole will remain very low, comparable to that of the progenitor star, because both essentially radiate at the Eddington luminosity.

These black holes would become detectable with JWST (or perhaps with future X-ray missions) only after significant mass growth. However, whether or not these black holes can grow sufficiently is being debated because of gravitational recoil following the merger of two halos leading to possible expulsion of the black holes from their host halo. Alternatively black holes might form directly from the direct collapse of a primordial gas cloud and they are likely not detectable directly with JWST (AB=33.5-35) until they have grown a factor 10 or so.

Further progress in this area will require improved theoretical understanding and observation of the luminosity function of mini-AGNs and its variation with redshift. Unfortunately, the JWST will be unable to detect the direct collapse black holes and thus it will make discrimination between models indirect and model dependent. On this basis we expect that this question will not be fully answered by JWST.

6. When and how did the first stars form?

Assuming our present theoretical understanding of Population III formation is correct and dark matter is in the form of CDM with the standard power spectrum (see, e.g., O'Shea & Norman 2006), the density of Population III stars will depend primarily on the effectiveness of the Lyman-Werner feedback. If this feedback is as effective as it appears from present investigations, the formation rate of Population III stars will be around 10^{-6} stars Mpc^{-3} yr^{-1}. Isolated Population III stars will also be relatively faint in the non-ionizing continuum (AB~38.5-40 at z=10-25, compared to AB~31 achievable in 10^5 s exposures by JWST), because most of their energy output is in the ionizing continuum (Bromm et al. 2001b, Tumlinson et al. 2003) which is efficiently absorbed by the IGM. Thus, single Population III stars will be impossible to detect directly with JWST. Their HII regions might be more easily detected (few 10^{-21} erg s^{-1} cm^{-2}), especially if magnified by gravitational lensing, but the detection remains challenging (NIRSpec can reach 3×10^{-19} erg s^{-1} cm^{-2} in 10^5 s) especially when coupled to the low density on the sky.

The best bet to identify single Population III stars is to search for the Pair Instability Supernovae (PISN) that they are supposed to produce depending on their mass (Heger & Woosley 2002). These objects are bright (AB~25-27) and may or may not be searchable by JWST depending on model details still not well determined. In a recent study the range of expected surface densities ranges from below 1 SN yr^{-1} deg^{-2} to 70 SN yr^{-1} deg^{-2} (Weinmann & Lilly 2005, Trenti & Stiavelli 2009) At the high end of this range they could be searched effectively with a wide shallow JWST survey. However, at the faint end they would need to be found by some other means, e.g., wide field near-IR imaging with a JDEM-style mission or perhaps from the ground. Alternatively they could perhaps be found with a Gamma Ray Burst mission if PISN are visible as GRBs (which we presently don't know).

Another difficulty of searching for PISN with JWST is their expected slowly varying light curve. If the light curve resembles that of SN 2006 gy (which may or may not be a PISN at low-z, Smith et al. 2008) variations of a magnitude or two would occur at high redshift on timescales comparable to the lifetime of JWST. Thus, finding these objects by other techniques and studying them spectroscopically with JWST may be the best solution.

An indirect method to study Population III stars is that of looking for the abundance anomalies predicted by PISN models. This could be done from the ground or by JWST for high redshift objects. The first galaxies in particular might show abundance anomalies if they have been enriched by Population III stars.

The detection of several PISN as a function of redshift would enable us to determine a Population III star formation history that could serve as a first comparison with models. Thus, it is likely that in the post-JWST era we will have improved understanding of Population III formation thanks to theoretical progress and a few observation results based on chemical tracers and PISN. However, we should expect this question not to be fully answered in this time frame.

It is worth noting that the models still need improvement to capture the later stages of Population III formation and predict an initial mass function more reliably than we can do now (see, e.g., Abel et al. 2000, Tumlinson et al. 2004, O'Shea et al. 2005). Recently the first complete simulation from "first principles" was completed but it models a relatively late forming object partly because of the small simulated volume (Yoshida et al. 2008). It is possible that objects like this will be actually prevented from forming because of the LW and chemical feedback (Furlanetto & Loeb 2005, Smith et al. 2009, Trenti & Stiavelli 2009). In any case clearly more effort needs to be spent on high dynamic range simulations and this will require more powerful computing resources which should become available in the future.

7. Summary

The above discussion suggests that two very difficult and important questions pertaining to the First light and reionization epoch will still need to be answered in the post-JWST era: i) When and how did the first stars form? And ii) When and how did the active galactic nuclei form? This should keep this field active and exciting even in the next decade. Conversely we expect that progress in our understanding of the first galaxies and reionization will be major and such that at this stage it would be hard to predict what further studies, if any, might be required.

References

Abel, T. et al. 2000, ApJ, 540, 39
Baltz, E. A., et al. 1998, ApJ, 493, L1
Begelman, M. C., et al. 2006, MNRAS, 370, 289
Bouwens, R. J., et al. 2007, ApJ, 670, 928
Bouwens, R.J., et al. 2008, ApJ, 686, 230
Bromm, V., et al. 2001a, MNRAS, 328, 969
Bromm, V., et al. 2001b, ApJ, 552, 464
Bunker, A. J., et al. 2004, MNRAS, 355, 374
Davidsen, A. F. et al. 1996, Nature, 380, 47
Fan et al. 2006, AJ, 132, 117
Furlanetto, S. R., Loeb, A. 2002, ApJ, 579, 1
Furlanetto, S. R., Loeb, A. 2005, ApJ, 634, 1
Gao, L. et al. 2005, MNRAS, 363, 379
Heger, A., Woosley, S. E., 2002, ApJ, 567, 532
Jakobsen, P., et al. 1994, Nature, 370, 35
Kashlinsky, A., et al. 2007, ApJ, 666, L1
Komatsu, E., et al. 2009, ApJS in press (see also arXiv:0803.0547)
Machacek, M. E., et al. 2003, MNRAS, 338, 273
Madau, P., Rees, M. J. 2002, ApJ, 542, L69
Malhotra, S., Rhoads, J. E., 2006, ApJ, 647, L95
Naoz, S., et al. 2006, MNRAS, 373, 98
Oesch, P., et al. 2009, ApJ, 690, 1350
O'Shea, B. W., et al. 2005, 628, L5
O'Shea, B. W., Norman, M. L. 2006, ApJ, 648, 31
O'Shea, B. W., Norman, M. L. 2008, ApJ, 673, 14
Ricotti, M., et al. 2001, ApJ, 560, 591
Ricotti, M., et al. 2002, ApJ, 575, 49
Ryan, R. E., et al. 2007, ApJ, 68, 839
Shaver, P. A., et al. 1999, AA, 345, 380
Shull, M., Venkatesan, A. 2008, ApJ, 685, 1
Smith, B. D. et al. 2009, ApJ, 691, 441
Smith, N., et al. 2008, ApJ, 686, 485
Stiavelli, M., et al. 2004, ApJ, 610, L1
Tegmark, M., et al. 1997, ApJ, 474, 1
Thompson, R. I., et al. 2007, ApJ, 666, 658
Trenti, M. Stiavelli, M. 2007, ApJ, 667, 38
Trenti, M., Stiavelli, M. 2008, ApJ, 676, 767
Trenti, M., Stiavelli, M. 2009, ApJ in press (see also arXiv:0901.0711)
Tumlinson, J. et al. 2003, ApJ, 584, 608
Tumlinson, J. et al. 2004, ApJ, 612, 602
Venkatesan, A., et al. 2003, ApJ, 584, 621
Volonteri, M. 2007, ApJ, 663, L5
Weinmann, S., Lilly, S. J., 2005, ApJ, 624, 526
Wise, J. H., Abel, T. 2008, ApJ, 685, 40
Yan, H., Windhorst, R. A. 2004, ApJ, 612, L93
Yan, H., et al. 2005, ApJ, 634, 109
Yoshida, N., et al. 2008, Science, 321, 669