feature article

The first stars, as seen by supercomputers

Tom Abel

Today's telescopes cannot look far enough into the cosmic past to observe the formation of primordial stars. If you want to see that process, you need sophisticated numerical simulations.

Tom Abel is a member of the Kavli Institute for Particle Astrophysics and Cosmology, an associate professor of physics at Stanford University, and an associate professor at SLAC, all in Stanford, California. He is also a visiting professor at the University of Heidelberg and the Heidelberg Institute for Theoretical Studies in Heidelberg, Germany.

Some 400 000 years after the Big Bang, the universe was cool enough that electrons and protons could come together to form neutral hydrogen. Through study of the cosmic microwave background radiation, cosmologists have gained an excellent understanding of the state of the universe at that "recombination" time. By the time the universe was 1 billion years old, it contained small galaxies, gamma-ray bursters, and even bright quasars likely powered by supermassive black holes as heavy as a billion suns. From observations of such objects, astronomers have gleaned more information about the early cosmos.

But most of the first billion years of cosmic history has been inaccessible to direct observation. Cosmologists call that early period the dark ages and speculate about the nature of the earliest objects in the universe and their role in shaping what was to follow. Did black holes or galaxies form first? Which was created earlier, high-mass or low-mass stars? In which stars were the first carbon and oxygen nuclei forged and distributed to intergalactic space, a necessary prerequisite to forming stars that resemble those in the Milky Way? Could some of those earliest stars still be around today? When were hydrogen atoms ionized, and where did the energy come from to do that? Those are just a few of the many pressing questions that cosmologists theorize about, partly in anticipation of a new generation of telescopes that will begin to provide answers within the next 10 years.[1]

Gas joins dark matter

From cosmic-microwave-background and other observations, cosmologists have pieced together a convincing and detailed picture of the early universe. Within a few minutes of the Big Bang, the universe had 76% of all its gas in the form of ionized hydrogen and 24% as ionized helium, an elemental composition that remained unchanged until the first stars formed. Also present were traces of deuterium, beryllium, and lithium, which together contributed less than 1/10 000 of the cosmic mass density. During that epoch, the distribution of material was almost perfectly uniform throughout space. The small deviations likely stemmed from quantum fluctuations that grew exponentially during the period of cosmic inflation.

After recombination, almost all of the hydrogen was neutral; only about 2 in 10 000 protons and electrons were unbound. Consequently, essentially all the radiation left over from early times traveled freely—that is, without scattering off electrons. The cosmic microwave background is a record of that radiation, which has since traveled for 13.7 billion years on straight trajectories to our detectors.

The small number of free electrons was sufficient that photon–electron Compton scattering coupled the temperature of the radiation to that of the gas. The heated electrons could collide with protons, or they could exchange positions with an electron in a neutral hydrogen atom. As a result, ion temperature and neutral-gas temperature were also coupled. Those mechanisms for thermally coupling radiation and the primordial gas were effective until the universe was approximately 10 million years old. From that time on, the expansion of the universe cooled the gas faster than Compton scattering could heat it. Consequently, the gas may have attained temperatures as low as a few kelvin before any structure formed in the universe. At such low temperatures, and with densities only around 100 particles per cubic meter, the gas pressure was tiny. Meanwhile, dark-matter objects were forming hierarchically; small ones collapsed first and continuously accreted and merged to make larger objects. The gas accumulated in the potential wells of dark-matter lumps as soon as the gravitational attraction to the lumps overwhelmed the gas's low pressure forces.

But additional processes must take place before stars and black holes can form in those dark-matter-dominated halos, which have masses of about 10 000 solar masses ($M_\odot$) and a characteristic gravitational potential energy (the technical term is "virial temperature") on the order of a few hundred kelvin. The trapped gas constantly rearranges itself, trying to attain a stable equilibrium by balancing its internal pressure forces against the inward gravitational attraction of the dark matter. As gas continues to fall in, accretion shocks form a few hundred light-years from the center of the lump (see figure 1a). Infalling gas, accelerated by gravitational force, turns its kinetic energy into internal, thermal energy as it passes those shocks. Lumps begin to merge, and as they do so, both the mass and the virial temperature of the resulting protogalactic object increase.

Eventually, the temperature becomes high enough that chemical reactions change the makeup of the gas. In that early stage, one of the most important chemical processes is

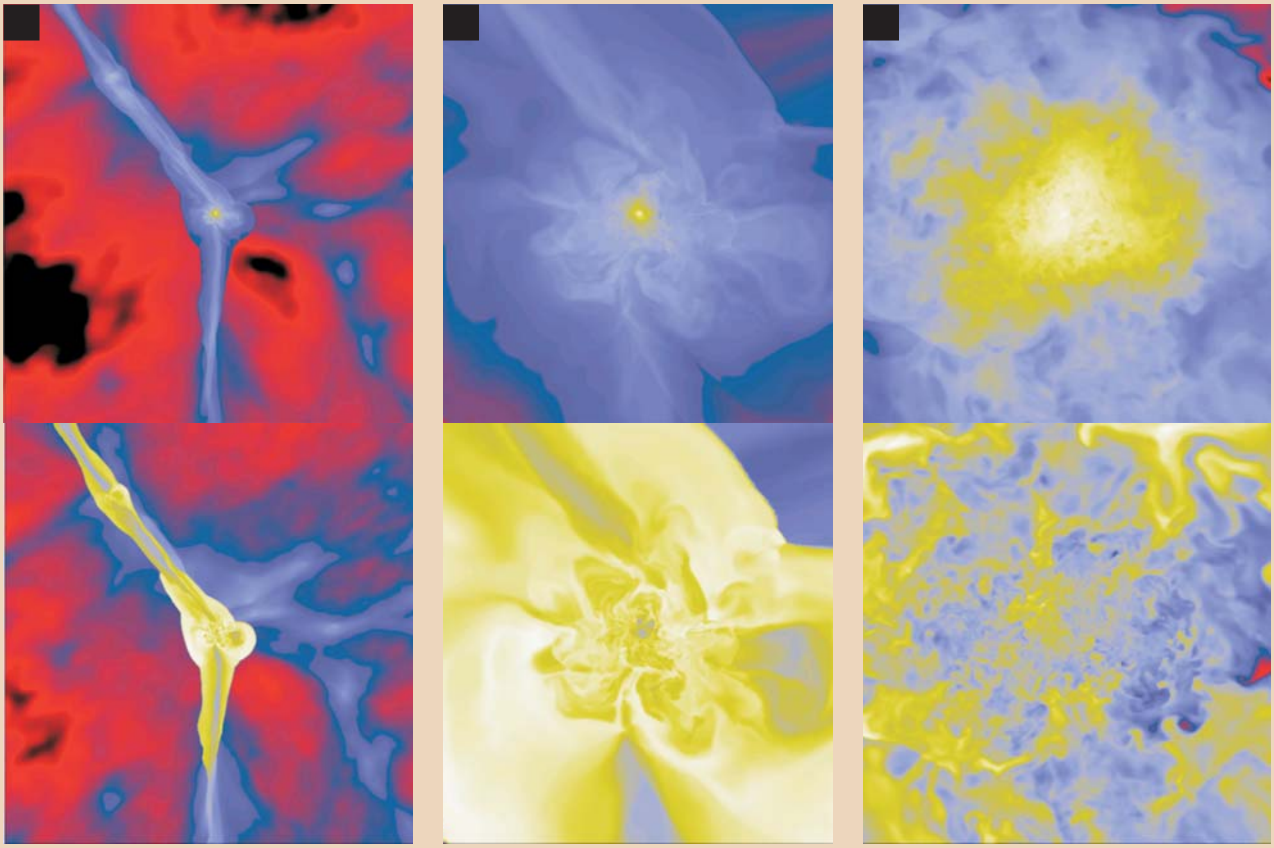

Figure 1. The gathering place. These six panels show density (top; red is less dense; yellow, denser) and temperature (bottom; red is 10 K; yellow, 1000 K) profiles of gas that falls into dark-matter gravitational potentials. **(a)** Visible here are spoke-like accretion shocks (blue on top; yellow on bottom) radiating from a gas-collecting lump of dark matter. Gas that subsequently falls in will convert kinetic energy to internal, thermal energy as it passes the shock. Pressure–volume work also heats up the gas. The regions shown are 10 000 light-years across. **(b)** Once the gas attains a high enough temperature, molecular hydrogen forms. The regions here are 1000 light-years across; yellow corresponds to a temperature of 1000 K. **(c)** Molecular decay cools the gas, which collects at the center; the area shown here is 15 light-years across and includes about 1000 solar masses of gas. Temperature ranges from 100 K to 1000 K. As the center of the concentrated molecular cloud contracts under its own gravity, it produces a dense region, at least two orders of magnitude smaller than shown here, that is amenable to star formation. (Simulation by Matthew Turk and Tom Abel; image by Tom Abel.)

catalyzed by the small fraction of free electrons left over from recombination, an event that took place some tens of millions to 100 million years earlier. The electrons join with neutral hydrogen to form the fragile negative hydrogen ion. The ion then rapidly attaches to a neutral hydrogen atom, forming a hydrogen molecule and ejecting the catalytic electron. Figure 1b depicts the dark-matter-and-gas object at that stage.

Even a single molecule per 1000 neutral atoms leads to a dramatic change in the thermodynamic behavior of the gas. That's because the lowest-energy rotational level of the hydrogen molecule has an excitation energy of only 512 K, low enough for the level (and higher ones, too) to become excited when the molecule collides with sufficiently fast neutral hydrogen atoms. With a decay time of hundreds of years, the diatomic molecule is a poor emitter. But it is also a poor absorber. Thus any photon released from the decay of a rotational or vibrational level of the molecule will exit from the protogalactic object. Since the energy of the escaping photon is taken from the motion of the colliding neutral hydrogen atom, the emission process carries away some of the gas's internal energy and cools it down. With cooling comes lower pressure. Gravitational forces further contract the gas, which collects at the center of the dark-matter-dominated structures (see figure 1c).

The contraction slows once the gas reaches a critical density of 10^5 particles/cm^3, which corresponds to a temperature of about 200 K. At that critical concentration, molecular hydrogen changes its cooling behavior. For gas below the critical density, every collision leads to the emission of a photon; the denser the gas becomes, the faster it cools. Once the critical density is surpassed, however, the rates for exciting rotational and vibrational levels in the molecule become greater than the rates for radiative decay. It is no longer the case that every collisional excitation yields a photon; rather, the photon emission per molecule tends to a constant value and the characteristic time scale to radiate the internal energy stays approximately constant.

Not only does the physics of H_2 determine a characteristic density and temperature for the cosmic gas, it also fixes characteristic length and mass scales. Here's how: The temperature sets the speed of sound, the speed at which changes in cloud pressure can be communicated. The density sets the gravitational collapse time of the region. The characteristic length, or Jeans length, is the distance for which the sound propagation time would equal the gravitational collapse time. It is named after the English physicist James Jeans (1877–1946), who derived it more rigorously via a stability analysis of the linearized equations of hydrodynamics, with

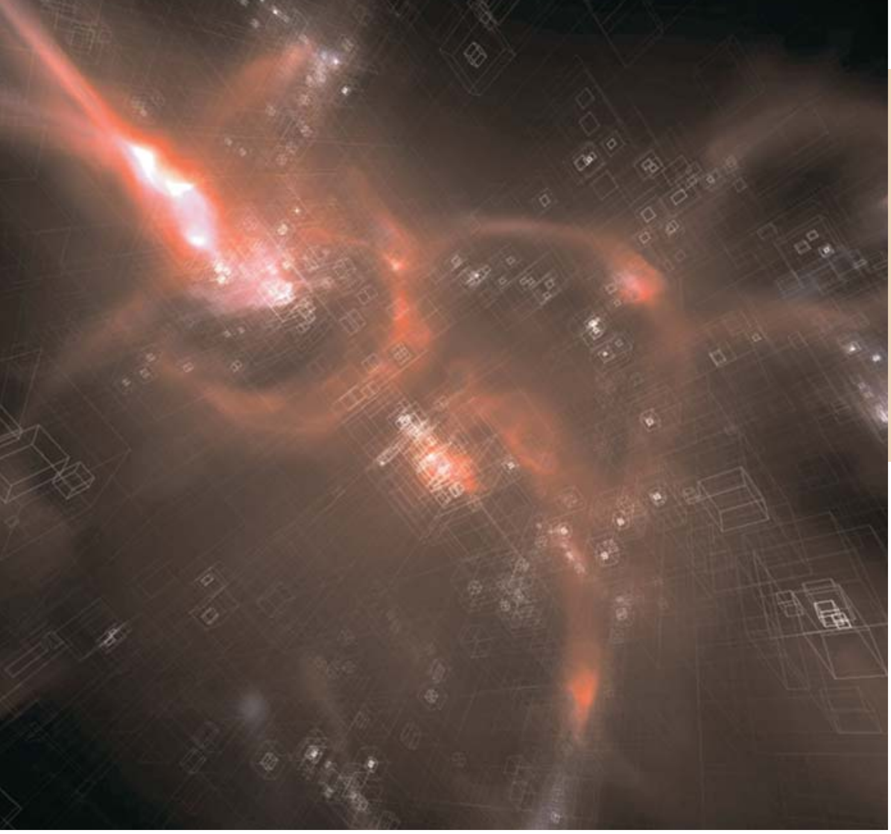

Figure 2. Evolving computational grids. With adaptive mesh refinement, a numerical simulation generates ever smaller grids to accurately capture the physics of contracting cosmic gas as it gets denser. The image here corresponds to a time 200 million years after the Big Bang and represents a region 2000 light-years across. The bright areas show where the cosmic gas density is greatest. Superimposed on the simulation frame is a hierarchy of grids. The smallest grids tend to track the bright regions of high density. (Simulation by John Wise and Tom Abel; image by Ralf Kaehler and Tom Abel.)

gravity included. When the scale of an object is larger than the Jeans length, gravity dominates over internal pressure forces. Length and density together determine a mass scale. The above considerations suggest that when the gas clumps contained in the centers of dark matter halos begin to contract due to their own gravity, the gas has a characteristic mass about 100 $M_\odot$.

A fully molecular cloud

Quite a bit more physics and chemistry becomes relevant as the central parts of the gas cloud contract further. Notably, once the density has reached about 10^8 particles/cm^3, three-body collisions of neutral hydrogen can render the collapsing material fully molecular. The binding energy released in the molecular formation is substantial, and most of it goes into heating the gas to temperatures above 1000 K.

The now fully molecular cloud starts to trap the radiation emitted by its constituent molecules. Consequently, it becomes increasingly difficult to cool the material by exciting rotational or vibrational levels and ejecting photons. Indeed, once the density rises to 10^{12} particles/cm^3, radiation emitted in the interior of the cloud may scatter and be converted into heat through collisional de-excitation between H_2 and hydrogen atoms or other H_2 molecules. The Jeans mass corresponding to the turning on of those processes is about 1 $M_\odot$. However, the same collisions that heat the gas also broaden the spectral lines and enable collision-induced emission with a much greater spectral coverage. That broadening substantially increases the cooling efficiency of the gas once its density has reached 10^{15} particles/cm^3, which corresponds to a Jeans mass of about 0.01 $M_\odot$. The gas contracts further and eventually gets optically thick to its own cooling radiation; by definition, it has formed the initial protostar. During that contraction phase, all the radiation produced originates from gravitational potential energy gained in the contraction. By now, roughly 100 million years after the Big Bang, typical time scales have dropped to minutes, but accretion continues for another 100 000 years, during which time the protostar swells to tens of solar masses.

Computational techniques

An understanding of the formation of the first luminous objects must involve a wealth of physics, including cosmic expansion, gravity, dark-matter dynamics, hydrodynamics, nonequilibrium chemistry, and radiative processes. The numerical simulations that incorporate all that physics are inherently three dimensional. Moreover, the simulations require an extraordinary dynamic range: The smallest spatial-resolution elements and time steps are minuscule compared to the size of the simulation volume and total time evolved in the calculation.

In the mid to late 1990s, Greg Bryan and Michael Norman, both then at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, implemented a numerical code called Enzo, the first cosmology code to use an adaptive mesh refinement technique.[2] The technique was not novel in and of itself—the engineering and aerodynamics literature of the 1980s includes discussions of adaptive meshes—but the application to first-star formation required a much enhanced dynamic range.[3] Simulations employing an adaptive mesh start with as large a computational box as possible and periodic boundary conditions. The computer stores such hydrodynamic quantities as density, internal energy per gram, velocities, and chemical compositions. Relevant equations are cast in a comoving form so that the mean cosmic expansion is absorbed in the coordinates; material is advected on the computational grid only once its motion starts to differ from the mean expansion of what is initially a nearly homogeneous and isotropic universe.

Once gravity induces some amount of matter to contract, the force will act to further concentrate the material. For dark matter, though, the gravitational collapse cannot proceed far because dark matter cannot dissipate its kinetic energy. As a result, a dark-matter lump reaches a dynamical equilibrium and collapse halts when the kinetic energy the lump gains from falling in its gravitational potential equals half the magnitude of the gravitational potential energy.

On the other hand, cosmic gas can collapse to scales as small as the Schwarzschild radius, smaller than the initial size by a factor of perhaps many trillions. Computationally, the collapse in length through many orders of magnitude is captured with "child" grids made of cells whose size is $1/n$ that of the surrounding parent grid. (Typically, $n = 2$.) Child grids beget further progeny to create an evolving hierarchy of grids optimally constructed to capture the physics of the underlying density field. Figure 2 shows a typical time step in a simulation, with a hierarchy of grids superposed.

Grids on a given level of the hierarchy do not overlap. Typically, calculations are run first for parent grids, which

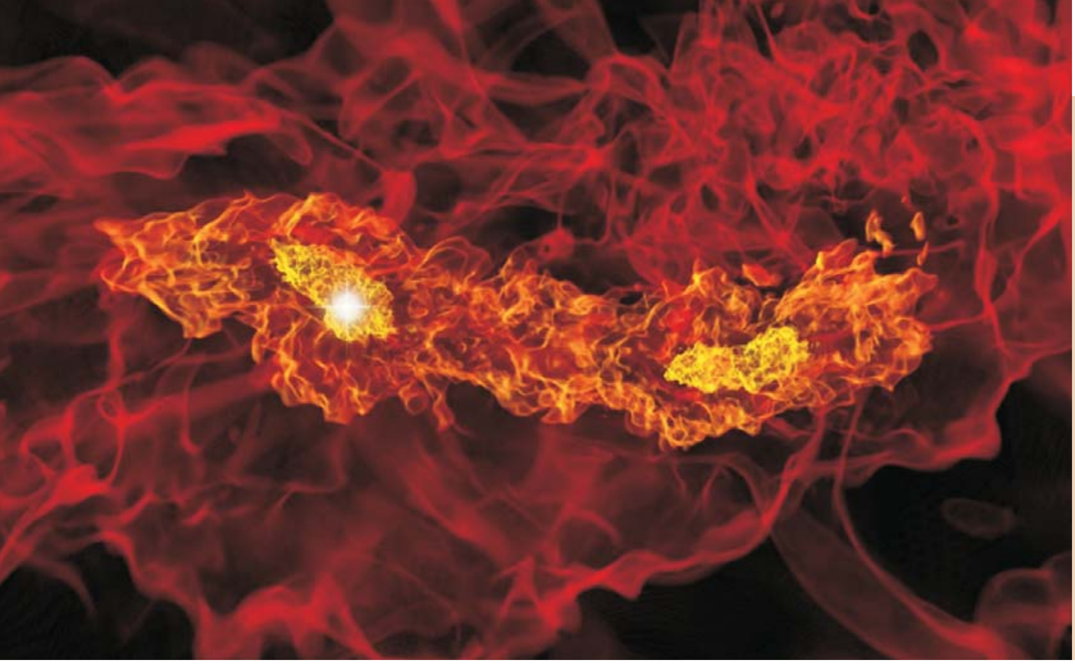

Figure 3. Neighboring protostars. The two star-forming regions in this simulation are separated by just a few hundred Earth–Sun distances. Shadings indicate density; brighter regions are more dense. The time is 200 million years after the Big Bang. The cover image is from the same simulation, but it presents a view from a different vantage point, from which only one star-forming region is evident. (Simulation by Matthew Turk, Brian O'Shea, and Tom Abel; image by Ralf Kaehler and Tom Abel.)

pass on their state variables to their child grids as boundary values. Once the boundary values are established, the simulation can advance the time step for all the grids on a given level. An important consideration is to account for fluxes across boundaries and, in particular, to ensure the conservation of mass, momentum, and energy moving between the parent and child grids.

More than motion

Unlike hydrodynamic interactions, the gravitational force is not local. For the largest of the grids, the associated Poisson equation can be efficiently solved with fast Fourier transform techniques. For the finer grids, fast multigrid approaches are best. To handle the dark-matter component, simulations use *N*-body methods, which have been around since the 1960s. They are convenient to implement and do an excellent job of respecting conservation laws. The idea is to represent the density field of dark matter by distinct, gravitationally interacting particles with well-defined position and velocity. The Poisson equation is then solved on the grid with the masses of both gas and dark matter as sources. The resulting gravitational potential is differentiated to yield the gravitational forces that accelerate the gas and dark matter. The calculated accelerations, in turn, determine new velocities for the particles and how they should be moved for the next time step. The result is a new density distribution, which leads to new accelerations, velocities, and positions.

There remains to be solved a set of ordinary differential equations that describe the chemical reactions in the gas. The chemistry is crucial because, as we have seen, molecular hydrogen that forms away from equilibrium dominates the radiative losses of the primordial gas. In principle, many hundreds of reactions could be relevant, but studies conducted over many years have concluded that 20 or so key reactions suffice to capture the abundances of the 12 species principally affecting the thermal history of the gas.[4] Those species are protons, electrons, neutral hydrogen atoms, helium and its two ionized forms, deuterium, ionized deuterium, HD, molecular hydrogen, singly ionized H_2, and the negative hydrogen ion H^-.

Much research has led to extraordinarily robust and accurate numerical packages that solve the ordinary differential equations describing the chemical reaction network. However, if large-scale numerical implementation were to be run with those routines, chemistry would dominate the computational cost of the simulation. Therefore, my colleagues and I have found it advantageous to construct routines that are optimized for our simulations of the first stars; with those routines, the chemistry takes up only a small fraction of the computational budget.[5]

The numerical program outlined above has been implemented in Enzo, a now much-expanded open-source version of the code originally developed by Bryan and Norman more than a decade ago.[2] Other techniques exist, notably the smoothed particle hydrodynamics approach.[6] So far, results obtained with various methods are in excellent agreement.

What the simulations teach

Modern numerical simulations can follow all the physics described above, up to the point when the first protostar forms. They reveal that the cold material within which stars will form is found in significant amounts only at the center of the dark-matter halos. The gas moves at approximately the speed of sound and is strongly concentrated toward the center; its density falls off a bit more steeply than $1/r^2$. A 100-$M_\odot$ region at the very center contracts faster than its surroundings. The earliest protostars, with masses of about 0.01 $M_\odot$, form at the centers of those relatively quickly contracting regions. The earliest protostars are tens of solar radii ($R_\odot$) across—down by more than 10 orders of magnitude from the radii of the dark-matter halos in which they form. Simulations show high rates of accretion, and spherically symmetric models convincingly demonstrate that within 100 years, protostars will grow to have a solar mass confined to a region some hundreds of solar radii across. In some cases, as shown in figure 3, simulations capture the formation of two- or three-star systems.

It seems likely that the accreting protostars will not be able to arrest the material flowing in from the rapidly collapsing central clump and that, consequently, the first stars to form will be very massive. However, many uncertainties exist about the accretion phase. A crucial difficulty confronting current numerical techniques has to do with the Courant time step—that is, the time it takes a sound wave to cross a small-spatial-resolution element. As the protostar gets hotter, sound speeds can rise to hundreds of kilometers per second. The Courant time step corresponding to a resolution of, say, 0.01 $R_\odot$ would be less than a minute. To follow 100 000 years of protostar evolution would thus require 100 billion time steps. That's not feasible with today's technology, never mind