ASTROPHYSICS AND COSMOLOGY

Editors
B Sinha
R K Moitra

Narosa

Astrophysics and Cosmology
Edited by B. Sinha and R.K. Moitra

3

The Saha Equation and Beyond

Jayant V. Narlikar
Inter-University Centre for Astronomy and Astrophysics
Ganeshkhind, Pune, India

1. Introduction

The seminal contribution of Meghnad Saha was his ionization equation, now well known as Saha's Ionization Equation. To say that it was an important work in astrophysics would be understating it: for, the subject of astrophysics really got going only as a result of the Saha equation. Let us begin with an examination of this assertion.

Till the second decade of this century the main observational handle on studies of stars had been their luminosities and spectra. While the luminosity could give a crude estimate of the star's distance using the inverse square law of illumination, the spectrum contained a lot more information.

For example, the continuum spectrum did, in the first approximation resemble the black body spectrum which was well known in those days. If the star was generating energy inside it and radiating it away, then it was in a state of equilibrium and provided the amount radiated was negligible compared to the total store of radiation being scattered within the star one expected the equilibrium state to resemble the black body state. This enabled the astronomer to estimate the star's surface temperature.

With surface temperatures of the order of 3000 K and above, it became clear that the matter at surface was not likely to be in a state of neutral gas. With large thermal motions it would be impossible for the atoms to retain all of their orbital electrons and so they would be ionized. How would the state of equilibrium be in such circumstances? The Saha equation answered this important question by giving the relation:

$$\frac{N_i N_e}{N} = \left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{B}{kT}\right) \tag{1}$$

Here the numbers N_e, N_i, N denote the number densities of free electrons, ions and neutral atoms at temperature T; B being the binding energy of the atom. The ratio of binding energy to temperature appears in the exponential form in this equation, thus underscoring its critical effect on the equilibrium abundances of these three component species.

In Saha's equation we therefore see the broad link between atomic physics, thermodynamics and observational astronomy. The appearance of the Boltzmann's constant k in Eq. 1, the atomic binding energy, and the temperature indicate this tripartite relationship. This was the beginning for astrophysics: it was here that a clue was made available to interpret the spectrum of a star including the strengths of the emission and absorption lines in it in terms of the ambient state of ionization of the stellar envelope and its temperature.

The spectroscopy of stars gives us a classification with respect to the lines observed. By identifying lines with the atoms of different kinds the astrophysicist can use the atomic details to work out the surface temperature of stars by using Saha's equation. Thus, spectral classification in the well-known Hertzsprung-Russell diagram gives us a temperature scale.

Surface conditions of the star serve as valuable boundary conditions for stellar models which seek to give details of the unseen stellar interior. The classic equations of Eddington [1] are differential equations which give the march of physical quantities like density, temperature, pressure, luminous flux etc. from the centre outwards. To solve them completely the boundary conditions at the surface are required. This explains why the Saha equation was such an important stimulus for the early astrophysics.

Purpose of this talk, however, is to emphasize the wide applicability of the Saha equation to astrophysics: for the general impression is that the equation has relevance to stellar scenarios only. I will select there scenarios to illustrate my point, all of them far removed from the stellar astrophysics. The first relates to the popular theory of the origin of the microwave background radiation in the universe, second to the theory of the origin of light nuclei in the early universe and third to the composition of particles in the very early universe.

2. The Microwave Background

Presently popular big bang framework of cosmology envisages the following sequence of events since the origin of the universe in a big explosion. In early stages the universe is very hot, with typical particles of matter moving relativistically, i.e., as photons. Such a phase is said to be radiation dominated. Even electrons and protons move with speeds close to that of light provided the universe has a temperature of about ten thousand billion. As the universe expands, it cools and first the protons become non-relativistic and then the electrons.

Standard texts in cosmology, e.g. [2] give the relevant relations describing when this happens. The later cooler epochs have the universe 'dust dominated'. That is, the universe is mainly made of matter that has negligible random motions with respect to the cosmological rest frame. Denoting the scale factor of the expanding universe by S, the simple rule is that the temperature of the radiation drops in inverse proportion to S.

As we shall see in the next two sections, there are two critical phases of the radiation dominated era. During the period 1–200 seconds after the big bang, the temperature of the universe dropped from about 10 billion degrees to a few hundred million degrees. This was when the synthesis of nuclei took place. This is the epoch of the 'early universe'. However, the composition of nucleons (protons and neutrons, pions etc. was determined during the so-called 'very early universe', which may take us from ~ 10^{-43} s to ~ 10^{-6} s. We will briefly discuss these epochs later. Here we consider the final stages of the radiation dominated universe.

The presence of nuclei, free protons and electrons did not, however, have much effect on the dynamics of the universe, which was still radiation-dominated. But, these particles, especially the lightest of them, the electrons, acted as scattering centres for the ambient radiation and kept it thermalized. The universe was therefore quite opaque to start with.

However, as the universe cooled, the Coulomb electrical attraction between the electron and proton began to assert itself. In detailed calculations performed by P.J.E. Peebles, the mixture of electrons and protons and of hydrogen atoms was studied at varying temperatures. Because of Coulomb attraction between the electron and the proton, the hydrogen atom has a certain binding energy B. The problem of determining the relative number densities of free electrons, free protons (that is, ions), and neutral H-atoms in thermal equilibrium is therefore analogous to that we considered earlier for stars. Only difference is that the setting is cosmological rather than stellar. Following Eq. 1 we arrive at the formula relating the number densities of electrons (N_e), proton ($N_p = N_e$), and H-atoms (N_H) at a given temperature T:

$$\frac{N_e^2}{N_H} = \left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{B}{kT}\right) \tag{2}$$

where m_e = electron mass. This equation is a particular case of *Sahā's ionization equation.*

Writing N_B for the total baryon number density, we may express the fraction of ionization by the ratio

$$x = \frac{N_e}{N_B}$$

Then, since $N_H = N_B - N_e$, we get from Eq. 2

$$\frac{x^2}{1-x} = \frac{1}{N_B}\left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{B}{kT}\right) \tag{3}$$

For the H-atom, B = 13.59 eV. Substituting for various quantities on the right-hand side of Eq. 3, we can solve for x as a function of T. The results show that x drops sharply from 1 to near zero in the temperature range of ~ 5000 K to 2500 K, depending on the value of N_B. For example, for x = 0.003 at T = 3000 K for the case where the baryon density at present is about 2×10^{-30} g cm^{-3}.

Hence, by this stage most of the fee electrons have been removed from the cosmological brew, and as a result the main agent responsible for the scattering of radiation disappears from the scene. The universe becomes effectively transparent to radiation. This is called the 'recombination epoch'.

Thus the Saha equation essentially fixes the epoch when radiation decoupled from matter. Subsequent to this epoch, the radiation cooled more or less undisturbed by whatever process went on with the formation of large scale structures of matter. The microwave background we see today should therefore carry signatures of the recombination epoch intact. This conclusion has remarkable observational consequences since it enables us to probe the early universe by looking at the microwave background.

3. Primordial Nucleosynthesis

Let us now go further back in time to the 1-200 second epoch when the universe was hot enough for nucleosynthesis to have taken place. Here we encounter the Saha equation in a different setting, with atomic binding replaced by nuclear binding. Free protons and neutrons can combine to form bigger and bigger nuclei provided their random speeds are slow enough for them to be trapped in the nuclear potential wells. The calculation which was first attempted by George Gamow in the late 1940s is described briefly as follows.

A typical nucleus Q is described by two quantities: A = atomic mass and Z = atomic number, and is written*

$$^{A}_{Z}Q$$

This nucleus has Z protons and $(A - Z)$ neutrons. If m_Q is mass of the nucleus, its binding energy is given by

$$B_Q = [Z\, m_p + (A - Z)m_n - m_Q]c^2 \tag{4}$$

Let us now consider a unit volume of cosmological medium containing N_N nucleons, bound or free. Since masses of protons and neutrons are nearly equal, we may denote the typical nucleon mass by m. Thus $m_n \approx m_p = m$. If there are N_n free neutrons and N_p free protons in the mixture

$$X_n = \frac{N_n}{N_N}, \quad X_p = \frac{N_p}{N_N} \tag{5}$$

will denote the fractions by weight of free neutrons and free protons. If a typical bound nucleus Q has atomic mass A and there are N_Q of them in our unit volume, we may denote the weight fraction of Q by

$$X_Q = \frac{N_Q\, A}{N_N} \tag{6}$$

Now at very high temperatures ($T \gg 10^{10}$ K), the nuclei are expected to be in thermal equilibrium. However, even at these temperatures the usual

*Sometimes the suffix Z is suppressed.

formulae for non-relativistic thermodynamics will apply. Further, since we are now concerned with relative number densities, we can no longer ignore the chemical potentials. Thus g_Q describing spin multiples

$$N_Q = g_Q \left(\frac{m_Q kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(\frac{\mu_Q - m_Q c^2}{kT}\right) \tag{7}$$

where we have introduced the chemical potentials μ_Q. Since chemical potentials are conserved in nuclear reactions

$$\mu_Q = Z\,\mu_p + (A - Z)\mu_n \tag{8}$$

assuming that the nuclei were built out of neutrons and protons by nuclear reactions.

Unknown chemical potentials can be eliminated between Eq. 7 and similar relations for N_p and N_n. Result is expressed in the form

$$X_Q = \frac{1}{2}\, g_Q\, A^{5/2}\, X_p^Z\, X_n^{A-Z}\, \xi^{A-1} \exp\left(\frac{B_Q}{kT}\right) \tag{9}$$

where

$$\xi = \frac{1}{2} N_N \left(\frac{mkT}{2\pi\hbar^2}\right)^{-3/2} \tag{10}$$

Note that Eq. 9 is a reincarnation of the Saha Eq. 1 with nuclear binding replacing the atomic Coulomb binding!

For an appreciable buildup of complex nuclei, T must drop to a low enough value to make exp (B_Q/kT) large enough to compensate for the smallness of ξ^{A-1}. This happens for nucleus Q when T has dropped down to

$$T_Q \sim \frac{B_Q}{k\,(A-1)\,|\ln \xi\,|} \tag{11}$$

Let us consider what happens when we apply the above formula to the nucleus of ^{4}He. The binding energy of this nucleus is $\cong 4.3 \times 10^{-5}$ erg. If we substitute this value in Eq. 11 and estimate N_N from the presently observed value of nucleon density of around 10^{-6} cm^{-3}, we find that T_Q is as low as $\sim 3 \times 10^9$ K. However, at this low temperature the number densities of participating nucleons are so low that four-body encounters leading to the formation of ^{4}He are extremely rare. Thus the underlying assumption of thermodynamic equilibrium (which requires frequent collisions) leading to Eq. 11 becomes invalid. We therefore need to proceed in a less ambitious fashion in order to describe the buildup of complex nuclei.

Hence we try using two-body collisions (which are not so rare) to describe the buildup of heavier nuclei. Thus deuterium (d), tritium (^{3}H), and helium

(^{3}He, ^{4}He) are formed via reactions like

$$\begin{aligned} p + n &\leftrightarrow d + \gamma \\ d + d &\leftrightarrow {}^3\text{He} + n \leftrightarrow {}^3\text{H} + p \\ {}^3\text{H} + d &\leftrightarrow {}^4\text{He} + n \end{aligned} \tag{12}$$

Since formation of deuterium involves only two-body collisions, it quickly reaches its equilibrium abundance as given by

$$X_\text{d} = \frac{3}{\sqrt{2}} X_\text{p} X_\text{n} \xi \exp\left(\frac{B_\text{d}}{kT}\right) \tag{13}$$

However, binding energy B_d of deuterium is low so that unless T drops to less than 10^9 K, X_d is not high enough to start further reactions leading to ^{3}He and ^{4}He. In fact reactions given in Eq. 12 with the exception of first one do not proceed fast enough until the temperature has dropped ~ 8×10^8 K.

Although at such temperatures nucleosynthesis does proceed rapidly enough, it cannot go beyond ^{4}He. This is because there are no stable nuclei with $A = 5$ or 8, and nuclei heavier than ^{4}He. So the process terminates there. Detailed calculations by several authors have now established this result quite firmly.

There is a fairly good agreement between these calculations and observational estimates of the light nuclei, agreement at least good enough to generate confidence in the big bang picture of a early hot universe. One could equally well argue that the success of the calculation generates confidence in the thermodynamic equilibrium picture conceived by Meghnad Saha.

4. The Very Early Universe

The successes of primordial nucleosynthesis calculation and observations of microwave background generated enough confidence amongst cosmologists of the 1970s and 1980s to enable them to tackle the more ambitious problems of the origin of subatomic particles. Here sequence of relationships is

Subatomic structure at smaller length scale
⇒ Matter at higher momenta (by the uncertainty principle)
⇒ Matter at higher energies (by relativity)
⇒ Matter at the earlier epochs of the universe

This last step is the outcome of big bang models. The solution of Einstein's equations of relativistic cosmology lead to a time-temperature ($t - T$) relationship after the big bang instant as

$$t = \left(\frac{3c^2}{16\pi Ga}\right)^{1/2} g^{-1/2} T^{-2} \tag{14}$$

This relation can be expressed as

$$t_{\text{second}} = 2.4\, g^{-1/2}\, T_{\text{MeV}}^{-2} = 2.4 \times 10^{-6}\, g^{-1/2}\, T_{\text{GeV}}^{-2} \quad (15)$$

Here G = gravitational constant, a = radiation constant and g = the statistical weightage to be attached to the particle distribution functions to include their different spin states. Typical particles are protons (p), neutrons (n), elec-trons (e), neutrinos (ν), pions (π), muons (μ), photons, (γ), etc. In thermo-dynamic equilibrium the bosons have different distribution functions from fermions, thus leading to different weight-factors. For bosons g is simply the number of spin states (e.g., $g_\gamma = 2$) while for fermions g is (7/8) times the number of spin states (e.g., $g_e = 7/4$).

If we go by the current wisdom, neutrons, protons and pions are bound states of quarks while the electrons, muons and neutrinos from the families of leptons. We can replace the calculations for bound nucleons of the previous section by those of bound quarks, for example. This will again bring out the essence of Saha's equation, viz. the relative ratios of various nucleons and mesons should in principle be determinable from the physics of the quark-gluon plasma phase.

This calculation is not so clearcut as the nucleosynthesis one because the basic physics is still somewhat speculative. The outcome of such a calculation would, however, be of great interest to the cosmologists since the neutron to proton ratio which is basic to the nucleosynthesis calculations should be determined by the above phase.

Going further back in history of the universe, one can do a calculation which determines the nucleon to photon ratio. For details of the pioneering calculations by Steigmen [3] we refer the reader to modern texts in cosmology (see for example [2]). The equilibrium ratio of a nucleon A to photon comes out to be

$$\frac{N_{\text{A}}}{N_\gamma} = \frac{\pi g^{1/2}}{7.2} \frac{x_*}{\zeta} \left(\frac{2Gm_{\text{A}}^2}{c\hbar} \right)^{1/2}, \ \zeta \approx 100 \quad (16)$$

Here $x_* = m_{\text{A}}c^2/kT$, $x_*/\zeta \sim 1/2$ and $g^{1/2} \sim 3$, so that the coefficient in front of the expression in parentheses is of the order unity. So the smallness of N_{A}/N_γ is directly related to the ratio of the strengths of the gravitational interaction and the strong interaction. Denoting this ratio by

$$\alpha_{\text{G}} = \frac{Gm_{\text{A}}^2}{c\hbar} \sim 6 \times 10^{-39} \quad (17)$$

we have

$$N_{\text{A}}/N_\gamma \sim \alpha_{\text{G}}^{1/2} \quad (18)$$

The strength of the electromagnetic interaction is measured by the fine structure constant $\alpha = e^2/\hbar c \sim 1/137$. Notice how weak the gravitational interaction is by comparison. Had G been considerably higher than it is, we could have ended with a larger value of N_{A}/N_ν.

This ratio more or less survives intact and poses a problem for cosmology. For, if we do an 'honest' calculation, $N_A = N_{\bar{A}}$, thus implying that both the particle A and its antiparticle should be equally abundant. Further, the number of photons per nucleon should be $\sim 10^{19}$. The observations give this number as $\sim 10^9$ and also observations do not indicate the predicted matter-antimatter balance on the Hubble distances scale. There is no direct evidence for large scale abundance of antimatter in the universe. We will not go into the proposed scenarios for resolving these difficulties: we give the above instances as examples of the extended applications of Saha's equation.

Conclusion

To conclude, these three examples from cosmology serve to emphasize the wide applicability of Saha's work today. When Saha was working on his equation cosmology was in its infancy. Therefore, Saha himself may not have imagined that his result would have important implications for cosmology which has come a long way both observationally and theoretically since Saha's days. However, work of a fundamental nature in physics inevitably finds unexpected and wider applications. The Saha equation is an ideal example of this dictum.

References

1. A.S. Eddington. *The Internal Constitution of the Stars*, Cambridge University Press, 1926.
2. J.V. Narlikar. *Introduction to Cosmology*, Cambridge University Press, 1993.
3. G. Steigman. *Annual Reviews of Astronomy and Astrophysics*, **14**, 339, 1976.