

Meghnad Saha's Ionization Equation (1920) and India's Entry into Astrophysics

Y C Kamala

Department Of Physics, The National College, Basavanagudi, Bangalore-560004, India

E-Mail: Yckamala@Gmail.Com

Abstract- In 1920 Meghnad Saha, a young lecturer at the University College of Science, Calcutta, published a theory of thermal ionisation that explained the sequence of stellar spectra as a sequence of temperatures. The Saha equation became a foundation of astrophysics and later of plasma physics, and it marked the entry of Indian science into theoretical astrophysics. This paper revisits the equation in its historical and physical context. It sets out the thermodynamic argument by which Saha obtained it, shows that the numerical form he published is consistent with the modern statistical-mechanical expression, and reproduces with the Saha and Boltzmann equations the calculations that made the theory persuasive: the degree of ionisation of hydrogen as a function of temperature and pressure, the maximum strength of the Balmer lines near 10 000 K, the successive ionisation of calcium, and the progression of line-forming populations of calcium, hydrogen and helium that underlies the spectral sequence from M to O stars. The pressure dependence of ionisation, which explains the differences between giant and dwarf stars, is also quantified. The paper then discusses how the theory was taken up by Fowler, Milne and Payne, how it relates to the earlier work of Eggert, and what it meant for science in India, where it was produced without access to a telescope.

Keywords: Meghnad Saha; ionization equation; stellar spectra; Boltzmann equation; Balmer lines; spectral classification; history of astrophysics in India

I. INTRODUCTION

By the end of the nineteenth century astronomers had classified the spectra of many thousands of stars into a sequence, now written O, B, A, F, G, K, M, in which the lines of hydrogen, helium, calcium and other elements rise and fall in strength. Whether this sequence reflected differences in chemical composition, temperature or something else was unclear [1]. The question was answered in 1920 and 1921 by Meghnad Saha (1893–1956), then a lecturer at the University College of Science in Calcutta. Saha showed that the atoms in a stellar atmosphere are ionised to an extent that depends on temperature and pressure, that the strength of a spectral line depends on the number of atoms in the appropriate state of ionisation and excitation, and that the spectral sequence is therefore primarily a sequence of temperatures [2], [3].

The result was important for astrophysics and for India. For astrophysics it supplied, for the first time, a physical theory of stellar spectra and opened the way to quantitative analysis of stellar atmospheres [4]–[6]. For India it was the first contribution by an Indian scientist to a central problem of modern astrophysics, made in a city without a research observatory, by a physicist trained in applied mathematics

[7], [8]. The early phase of modern physics in India, of which Saha's work is a highlight, has also been examined in a Kannada-language doctoral analysis [9].

This paper has three aims. The first is to set out the physical reasoning of Saha's equation, including the numerical form in which he published it. The second is to reproduce, with modern constants and atomic data, the calculations that show how the equation explains the spectral sequence. The third is to assess what the episode meant for India's place in astrophysics.

II. HISTORICAL BACKGROUND

Two developments made Saha's theory possible. The first was Bohr's model of the atom of 1913, which gave a physical meaning to ionisation energies and excitation levels and explained the series of hydrogen lines [10]. The second was the application of chemical thermodynamics to the dissociation of atoms. In 1919 John Eggert used the Nernst heat theorem to estimate the degree of ionisation of iron in the hot interiors of stars [11]. Saha, who had been teaching thermodynamics and the new atomic physics to postgraduate

students in Calcutta, saw that the same reasoning could be applied to the atmospheres of stars, where the lines are formed [7], [12].

Saha's first paper, "Ionization in the solar chromosphere", appeared in the *Philosophical Magazine* in 1920 [2]. It treated ionisation as a reversible reaction, $\text{Ca} \rightleftharpoons \text{Ca}^+ + e^-$, applied the law of mass action, and used the result to explain why the lines of ionised calcium extend to great heights in the solar chromosphere. In the following year, during a stay in Europe, he extended the theory to the full sequence of stellar spectra in "On a physical theory of stellar spectra", published in the *Proceedings of the Royal Society* [3]. The essential claim was that each spectral class corresponds to a temperature at which particular ionisation stages and excitation levels are populated

The theory was quickly taken up in Britain and the United States. Fowler and Milne refined it by using the temperatures at which lines reach their maximum strength rather than those at which they first appear, and they showed that the pressures in stellar atmospheres are much lower than Saha had assumed [4], [13]. Cecilia Payne used the refined theory to derive the relative abundances of the elements in stars and found that hydrogen and helium are far more abundant than had been thought [5]. Russell and his collaborators made the theory central to the analysis of stellar spectra in the 1920s [6]. DeVorkin has examined why Saha himself played a limited part in this subsequent development, despite having founded the theory [14].

III. THEORY

1. Excitation

In thermal equilibrium at temperature T , the numbers of atoms in two energy levels a and b of the same ionisation stage are related by the Boltzmann equation

$$\frac{N_b}{N_a} = \frac{g_b}{g_a} \exp\left(-\frac{E_b - E_a}{k_B T}\right)$$

where g is the statistical weight and E the excitation energy. The total population of a stage is distributed over its levels according to the partition function

$$U(T) = \sum_i g_i \exp\left(-\frac{E_i}{k_B T}\right)$$

2. Ionisation as a chemical equilibrium

Saha treated ionisation as a reaction $A \rightleftharpoons A^+ + e^-$ in a mixture of ideal gases. For a pure gas of atoms of which a fraction x is ionised at total pressure P , the partial pressures are $p_+ = p_e = xP/(1+x)$ and $p_0 = (1-x)P/(1+x)$, so that the equilibrium constant is

$$K_p = \frac{p_+ p_e}{p_0} = \frac{x^2}{1-x^2} P$$

Thermodynamics gives the temperature dependence of K_p through the van 't Hoff relation,

$$\frac{d \ln K_p}{dT} = \frac{\Delta H}{RT^2}, \quad \Delta H = U + \frac{5}{2} RT$$

where U is the ionisation energy per mole at absolute zero and the term $\frac{5}{2} RT$ is the enthalpy carried by the free electron, treated as a monatomic ideal gas; the heat capacities of the atom and the ion cancel. Integration gives $\ln K_p = -U/RT + \frac{5}{2} \ln T + i$, and the Nernst heat theorem fixes the constant of integration i through the "chemical constant" of the electron gas. Converted to common logarithms, the result in the form that Saha published is [2], [3]

$$\log_{10} \left(\frac{x^2}{1-x^2} P \right) = -\frac{U}{4.571 T} + \frac{5}{2} \log_{10} T - 6.5$$

with P in atmospheres and the ionisation energy U in calories per mole. The factor 4.571 is $R \ln 10$ in calories per mole per kelvin.

3. The modern form

Statistical mechanics gives the same law with the statistical weights made explicit [15], [16]. For successive ionisation stages i and $i+1$, with number densities N_i and N_{i+1} and electron density n_e ,

$$\frac{N_{i+1} n_e}{N_i} = \frac{2 U_{i+1}}{U_i} \left(\frac{2 \pi m_e k_B T}{h^2} \right)^{3/2} \exp\left(-\frac{\chi_i}{k_B T}\right)$$

where χ_i is the ionisation energy and the factor 2 is the statistical weight of the free electron. With the electron pressure $P_e = n_e k_B T$, this becomes

$$\frac{N_{i+1}}{N_i} = \frac{2 k_B T U_{i+1}}{P_e U_i} \left(\frac{2 \pi m_e k_B T}{h^2} \right)^{3/2} \exp\left(-\frac{\chi_i}{k_B T}\right)$$

The bracketed factor is the inverse cube of the thermal de Broglie wavelength of the electron, $\lambda_{th} = h/\sqrt{2 \pi m_e k_B T}$. Equation (6) therefore has a simple interpretation: ionisation is favoured when each free electron has, on average, a large volume $1/n_e$ compared with λ_{th}^3 , which happens at high temperature and low density. Because $1/(n_e \lambda_{th}^3)$ is a very large number in stellar atmospheres, ionisation sets in at temperatures for which $k_B T$ is still only a small fraction of χ_i , typically between one tenth and one twentieth. This is why hydrogen, with an ionisation energy of 13.6 eV, is largely ionised near 10 000 K, where $k_B T$ is under 1 eV. For hydrogen-like statistical weights, $2U_{i+1}/U_i = 1$, Eq. (6) reduces to Eq. (5) with the constant

$$\frac{(2 \pi m_e)^{3/2} k_B^{5/2}}{h^3} = 3.33 \times 10^{-2} \text{ Pa K}^{-5/2}, \quad \log_{10} \frac{C}{1 \text{ atm}} = -6.48$$

Evaluated with present-day constants, the constant is -6.48 , against the -6.5 in Saha's equation, and $R \ln 10 = 4.576 \text{ cal mol}^{-1} \text{ K}^{-1}$, against his 4.571. Saha's numerical form is

therefore consistent with the modern expression to within the precision of the constants available in 1920.

4. Line strengths

The strength of an absorption line depends, to a first approximation, on the number of absorbers in its lower level. Combining Eqs. (1), (2) and (7), the fraction of all atoms of an element that are in level n of stage i is

$$\frac{N_{i,n}}{N_{\text{tot}}} = \frac{g_{i,n} e^{-E_{i,n}/k_B T}}{U_i(T)} \cdot \frac{N_i}{\sum_j N_j}$$

The first factor rises with temperature as excitation increases; the second eventually falls as the stage is ionised away. Their product has a maximum, and the temperature of that maximum, given by

$$\frac{\partial}{\partial T} \left(\frac{N_{i,n}}{N_{\text{tot}}} \right)_{P_e} = 0$$

is the basis of the method of maxima introduced by Fowler and Milne [4], [13].

IV. DATA AND METHOD

Ionisation energies were taken from Moore's compilation [17]: 13.598 eV for H I, 24.587 and 54.418 eV for He I and He II, and 6.113 and 11.872 eV for Ca I and Ca II. Partition functions were evaluated from Eq. (2) using the lowest levels of each species, which dominate at the temperatures considered (Table 1). Electron pressures of 2, 20 and 200 Pa span the range typical of the line-forming layers of stellar photospheres [18], [19]. Calculations were made at temperatures from 3000 to 50 000 K. The equations and constants are those stated above; no adjustable parameters were used.

Table 1. Atomic data used in the calculations [17].

Species	Ionisation energy (eV)	Levels included in U (energy in eV, statistical weight)	Line-forming level used
H I	13.598	0 (2); 10.20 (8); 12.09 (18)	$n = 2$ (Balmer)
He I	24.587	0 (1); 19.82 (3); 20.62 (1); 20.96 (9); 21.22 (3)	$2^3 P$ (447.1 nm)
He II	54.418	0 (2); 40.81 (8); 48.37 (18)	$n = 3$ (468.6 nm)
Ca I	6.113 (10)	0 (1); 1.88–1.90 (1, 3, 5); 2.52 (15); 2.71 (5); 2.93 (3)	ground (422.7 nm)
Ca II	11.872	0 (2); 1.69–1.70 (4, 6); 3.12–3.15 (2, 4)	ground (K line, 393.4 nm)

V. RESULTS

1. Ionisation of hydrogen

Fig. 1 shows the degree of ionisation of a pure hydrogen gas computed from Eq. (5) at total pressures of 10^{-4} , 10^{-2} and 1 atm. Ionisation rises steeply over a narrow range of temperature, and the temperature at which half the atoms are ionised increases with pressure: 8 730 K at 10^{-4} atm, 11 200 K at 10^{-2} atm and 15 360 K at 1 atm. Raising the pressure by four orders of magnitude raises the half-ionisation temperature by three quarters. This pressure dependence was the key to the later recognition by Fowler and Milne that stellar atmospheres are tenuous [4].

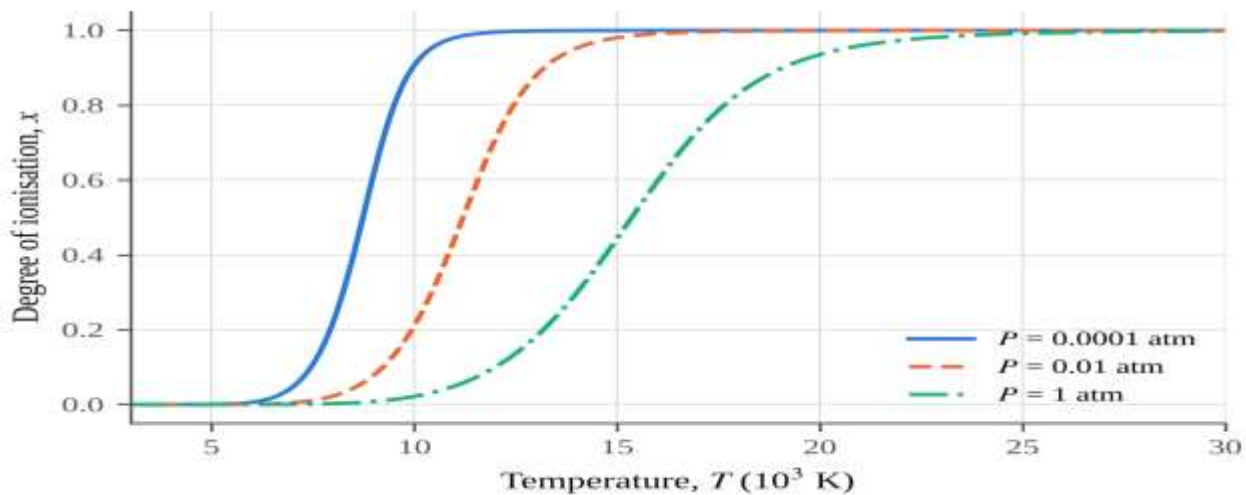


Fig. 1. Degree of ionisation of pure hydrogen as a function of temperature at three total pressures, from Saha's equation in the form of Eq. (5).

2. The Balmer maximum

The Balmer lines of hydrogen arise from the level $n = 2$, which lies 10.2 eV above the ground state. Fig. 2 shows the fraction of all hydrogen atoms in that level. At $P_e = 20$ Pa the fraction peaks at 9 870 K, at a value of only 8.7×10^{-6} ; at 2 Pa and 200 Pa the peaks occur at 8 780 K and 11 230 K. The Balmer lines are strongest in the A stars, whose

temperatures are close to 10 000 K. Below this temperature too few atoms are excited; above it too many are ionised. This single calculation explains the most prominent feature of the spectral sequence. The value at 20 Pa agrees with the standard textbook calculation for the same electron pressure, which places the maximum near 9 900 K [19], and it confirms that the implementation reproduces the established result.

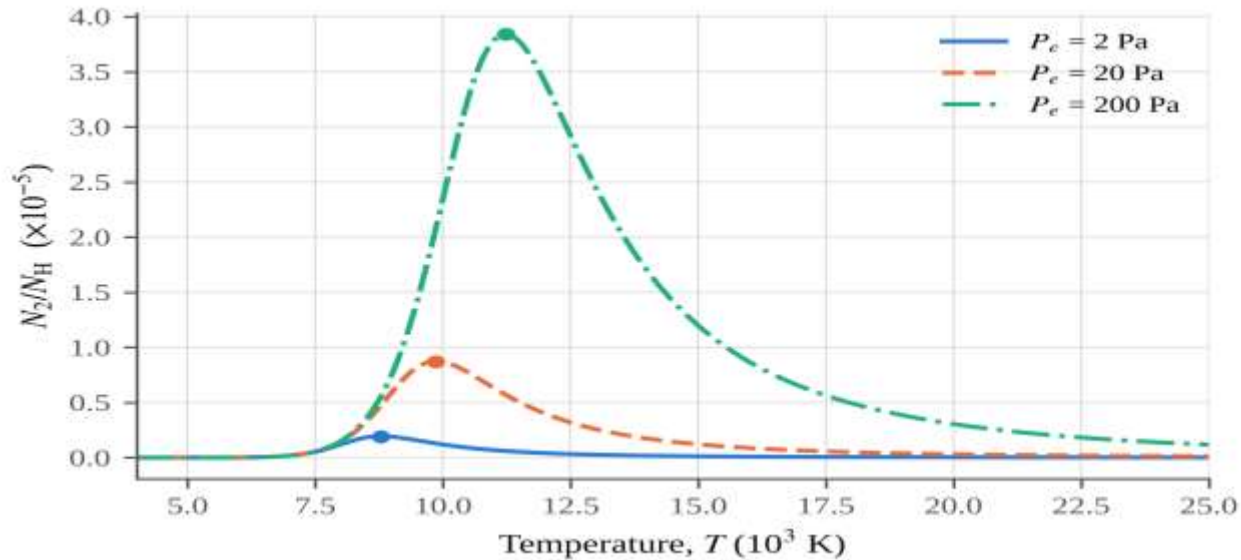


Fig. 2. Fraction of hydrogen atoms in the level $n = 2$ as a function of temperature at three electron pressures. Dots mark the maxima.

3. Successive ionisation of calcium

Fig. 3 shows the distribution of calcium among its first three ionisation stages at $P_e = 20$ Pa. Neutral and singly ionised calcium are equally abundant at 4 430 K, and singly and doubly ionised calcium at 8 690 K. Between these temperatures almost all calcium is singly ionised, with a maximum fraction of 0.994 at 6 410 K. This explains why

the H and K lines of ionised calcium dominate the spectra of the Sun and other G and K stars, and why they weaken in the hotter A stars even though calcium is no less abundant there. It was precisely the behaviour of these lines in the solar chromosphere that Saha's first paper addressed [2].

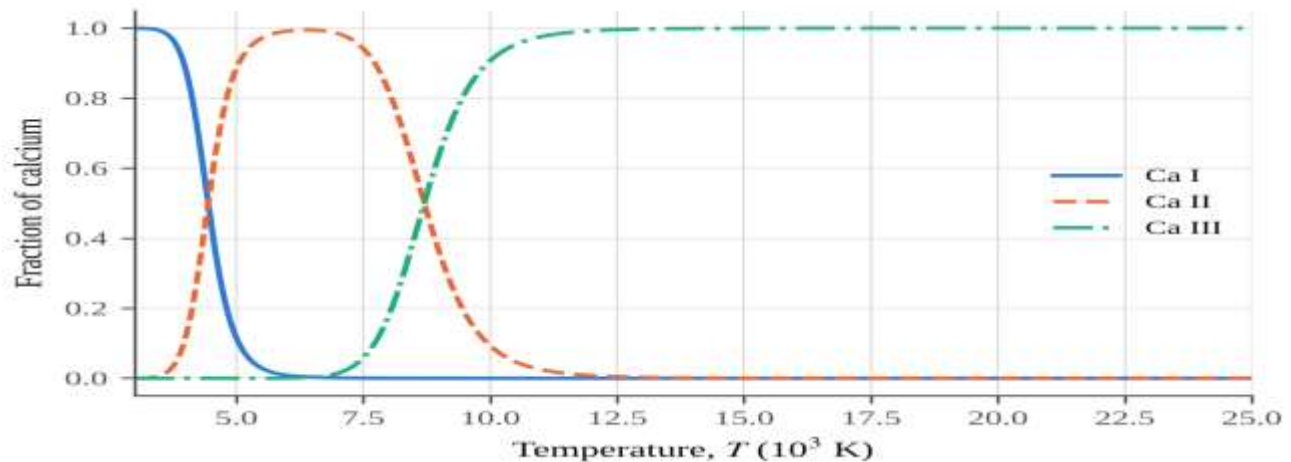


Fig. 3. Fractions of calcium in the stages Ca I, Ca II and Ca III as functions of temperature at $P_e = 20$ Pa.

4. The spectral sequence

Fig. 4 brings together the line-forming populations of Eq. (9) for five lines that mark the spectral sequence, each normalised to its maximum. At $P_e = 20$ Pa the population responsible for the neutral calcium line at 422.7 nm is largest below about 4 000 K, the ground state of Ca II peaks at 5 470 K, the $n = 2$ level of hydrogen at 9 870 K, the 2^3P level of

He I at 15 800 K, and the $n = 3$ level of He II at 34 300 K. The order of these maxima is the order of the spectral classes M, K–G, A, B and O. Saha's central insight is thus reproduced quantitatively: a single sequence of temperatures, acting through ionisation and excitation, produces the observed sequence of spectra without any change in composition.

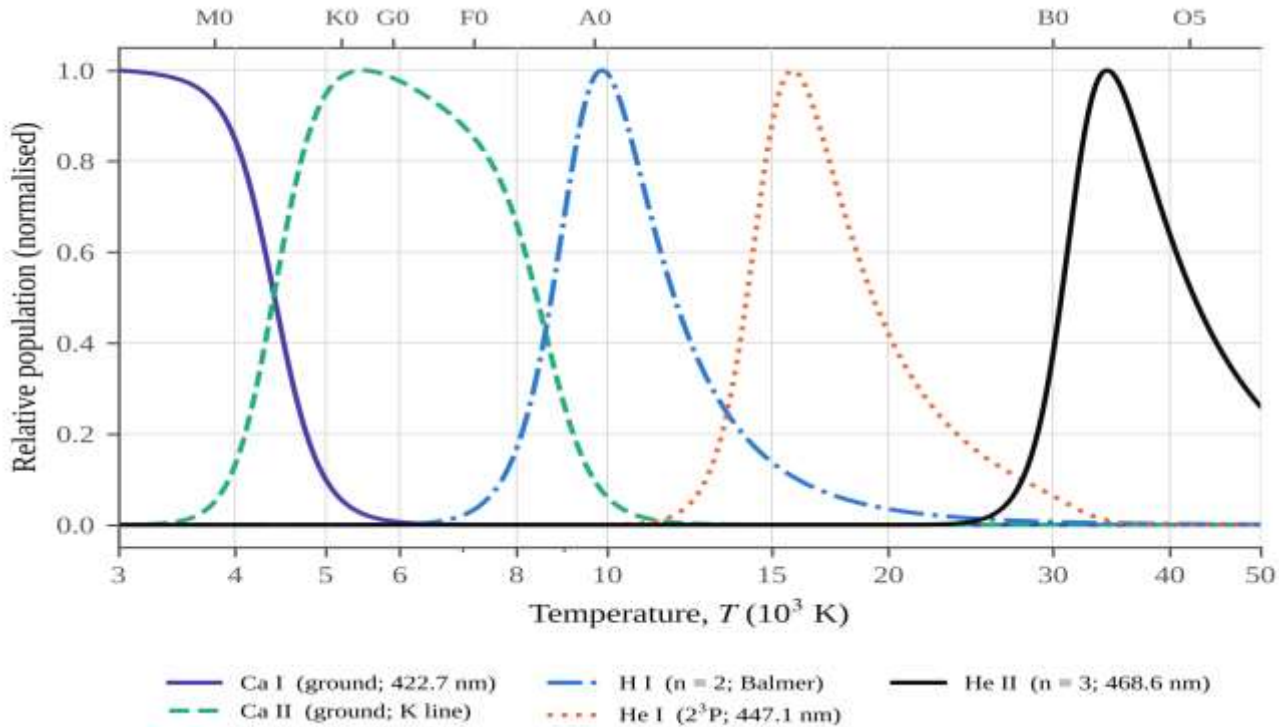


Fig. 4. Normalised populations of the lower levels of five diagnostic lines as functions of temperature at $P_e = 20$ Pa. The upper axis shows approximate effective temperatures of spectral classes.

The normalised curves for the ground state of Ca II and the $n = 2$ level of hydrogen cross near 8 600 K, but their absolute values tell a further story. At 6 000 K and $P_e = 20$ Pa the number of Ca II ions in the ground state per calcium atom exceeds the number of hydrogen atoms in the level $n = 2$ per hydrogen atom by a factor of 7.6×10^7 . With a calcium-to-hydrogen abundance ratio of about 2×10^{-6} , absorbers of the Ca II K line still outnumber absorbers of the Balmer lines by more than a hundred to one. This is why the H and K lines of ionised calcium are the strongest lines in the solar spectrum although hydrogen is hundreds of thousands of times more abundant than calcium, the very inference that allowed Payne to show that hydrogen dominates the composition of stars [5], [19].

The temperatures in Fig. 4 should not be identified exactly with the effective temperatures of the corresponding classes. The electron pressure varies along the main sequence, and

line strengths depend on opacity and on the structure of the atmosphere as well as on level populations [16], [18]. The

calculation nevertheless reproduces the order and approximate positions of the maxima, which is what Saha's theory claimed.

5. The pressure effect

Fig. 5 shows the temperature at which adjacent ionisation stages are equally populated, as a function of electron pressure. For hydrogen it rises from 8 250 K at 1 Pa to 10 430 K at 100 Pa; for the ionisation of Ca I from 3 810 to 4 870 K; and for Ca II from 7 420 to 9 580 K. A giant star, whose atmosphere has a much lower pressure than that of a dwarf of the same temperature, is therefore more highly ionised. Lines of ionised atoms are relatively stronger in giants, an effect used to separate giants from dwarfs spectroscopically. Saha recognised that pressure enters his equation alongside temperature [3], and the effect became a basis of spectroscopic luminosity classification [1].

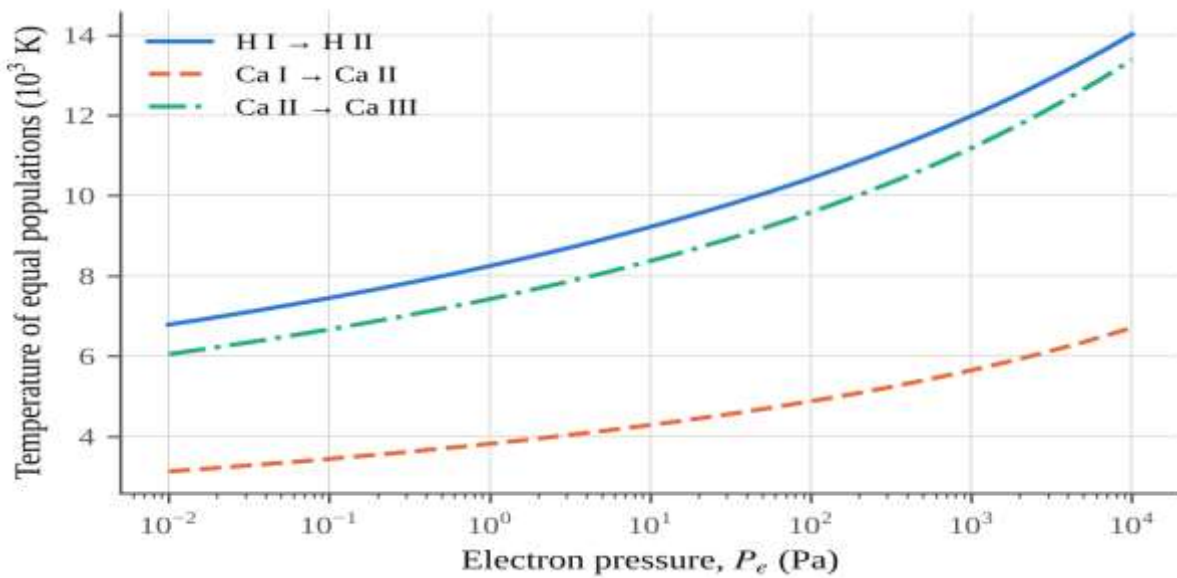


Fig. 5. Temperature at which adjacent ionisation stages are equally populated, as a function of electron pressure, for hydrogen and calcium

6. Marginal appearance and maxima

Saha estimated temperatures from the “marginal appearance” of lines, the temperature at which a line first becomes visible or finally disappears along the spectral sequence [3]. That criterion depends on the abundance of the element and on the sensitivity of the observations, since a more abundant element produces a visible line at a smaller fractional population. Fowler and Milne replaced it with the temperature at which a line reaches its maximum strength [4], [13]. Equation (10) shows why this is better: multiplying the population curve by a constant abundance does not change the position of its maximum. The maxima therefore yield temperatures that are independent of abundance, and, once the temperature scale is fixed, the observed positions of the maxima can be used to infer the pressure.

The calculations here show how sensitive the maxima are to pressure. The Balmer maximum moves from 8 780 K to 9 870 K to 11 230 K as the electron pressure rises by successive factors of ten (Section 5.2), about 1 100–1 400 K per decade. Using Saha’s assumption of relatively high pressures, a given maximum would be assigned too high a temperature. Fowler and Milne turned this sensitivity to advantage and concluded that the pressures in the reversing layers of stars are of the order of 10^{-4} atm [4]. Payne then used the corrected temperature scale to derive abundances, which required knowing the fraction of each element in the relevant level, exactly the quantity given by Eq. (9) [5].

Table 2 summarises the characteristic temperatures obtained.

Table 2. Characteristic temperatures computed from the Saha and Boltzmann equations.

Quantity	Condition	Temperature (K)
Half-ionisation of pure H	$P = 10^{-4} / 10^{-2} / 1 \text{ atm}$	8 730 / 11 200 / 15 360
Maximum of H $n = 2$ population	$P_e = 2 / 20 / 200 \text{ Pa}$	8 780 / 9 870 / 11 230
Ca I = Ca II	$P_e = 20 \text{ Pa}$	4 430
Ca II = Ca III	$P_e = 20 \text{ Pa}$	8 690
Maximum of Ca II ground population	$P_e = 20 \text{ Pa}$	5 470
Maximum of He I 2^3P population	$P_e = 20 \text{ Pa}$	15 800
Maximum of He II $n = 3$ population	$P_e = 20 \text{ Pa}$	34 300

VI. DISCUSSION: INDIA’S ENTRY INTO ASTROPHYSICS

Before 1920, research in astronomy in India was largely the work of British observers in government institutions. The Kodaikanal Observatory, where John Evershed discovered in 1909 the radial flow of gas in sunspots that bears his name, was the most important centre of solar research [8], [20]. Indian participation in astronomy was limited, and theoretical astrophysics hardly existed as a field anywhere.

Saha's contribution came from a different direction. He was not an astronomer and had no telescope. His resources were a university library, a training in mathematics and thermodynamics, and familiarity with the new atomic physics that he had acquired in preparing postgraduate lectures [7], [12]. The theory he produced was nevertheless decisive for observational astronomy, because it converted the qualitative sequence of spectral classes into a quantitative scale of temperature and pressure. Fowler, Milne, Payne and Russell built the subsequent research programme on it [4]–[6].

The circumstances of the discovery deserve emphasis. The University College of Science in Calcutta had been established only in 1914, and Saha and his contemporaries were among its first lecturers. They were required to teach subjects, such as thermodynamics, relativity and the quantum theory of the atom, that were new even in Europe, and they learned them largely from journals as they taught [7], [9]. Saha's theory grew directly out of this teaching. It brought together results from physical chemistry, atomic physics and spectroscopy that were rarely combined, and its originality lay as much in the combination as in any single step. Three features of the episode are significant for the history of science in India. First, it showed that original work of the first rank could be done in an Indian university in a field that was not tied to local data or to colonial administrative needs. Second, it illustrated the value of the new postgraduate departments in Calcutta, where young lecturers were expected to teach the latest physics and thereby came into contact with its open problems. Third, it revealed the limits of theory without observational resources.

DeVorkin's account of "Saha's fate" shows how the development of the theory passed to groups with access to observatories and spectra [14]. Saha himself turned to other fields, including the physics of the ionosphere and nuclear physics, and became a leading advocate of national planning for science [7], [21]. The equation also outlived its astronomical origins. In the laboratory, Langmuir and Kingdon used the closely related surface-ionisation law to explain the emission of ions from hot metal surfaces in alkali vapours [22], and the Saha equation remains a basic tool in plasma physics, used wherever the ionisation of a gas in thermal equilibrium must be estimated, from arcs and flames to fusion plasmas [23]. Its persistence across fields is a measure of the generality of Saha's insight.

VII. LIMITATIONS OF THE CALCULATIONS

The calculations assume local thermodynamic equilibrium, a single temperature and a fixed electron pressure for each

curve. They use partition functions truncated to the lowest levels, which is adequate at the temperatures shown but underestimates the partition functions of neutral atoms at the highest temperatures. Extending the hydrogen sum to $n = 10$ changes the partition function by only 0.007% at 10 000 K and by 1.2% at 15 000 K. At higher temperatures, however, the unrestricted sum grows without limit, a well-known difficulty of the isolated-atom partition function that is resolved in a real gas by the interactions that lower the ionisation energy and remove the highest levels [16]. The calculations also compute level populations, not line strengths, which depend in addition on oscillator strengths, opacity and the temperature structure of the atmosphere [16], [18]. These simplifications were also, in large part, those of Saha's original papers; the purpose here is to reproduce the logic of his theory rather than to construct a model atmosphere.

VII. CONCLUSION

Saha's ionisation equation, obtained in Calcutta in 1920 by applying chemical thermodynamics to atoms in stellar atmospheres, explained the sequence of stellar spectra as a sequence of temperatures. The numerical form he published agrees with the modern statistical expression to within the precision of the constants of his time. Calculations with the Saha and Boltzmann equations reproduce the Balmer maximum near 10 000 K, the dominance of ionised calcium in cooler stars, the ordering of hydrogen, helium and calcium lines through the spectral classes, and the pressure effect that distinguishes giants from dwarfs. The theory marked India's entry into astrophysics and demonstrated that fundamental work could be done in an Indian university, while the subsequent history showed how much the further development of such work depended on access to observational resources.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest: The author declares no conflict of interest.

Data availability: All data needed to reproduce the results are contained in the manuscript; the parameter values and equations used for every figure are stated in the text.

Ethics approval: Not applicable.

Author contribution: The sole author conceived the study, carried out the analysis and wrote the manuscript

REFERENCES

1. J. B. Hearnshaw, *The Analysis of Starlight: One Hundred and Fifty Years of Astronomical*

Spectroscopy. Cambridge, U.K.: Cambridge Univ. Press, 1986.

2. M. N. Saha, "Ionization in the solar chromosphere," *Philos. Mag.*, ser. 6, vol. 40, no. 238, pp. 472–488, 1920, doi: 10.1080/14786441008636148.
3. M. N. Saha, "On a physical theory of stellar spectra," *Proc. R. Soc. Lond. A*, vol. 99, no. 697, pp. 135–153, 1921, doi: 10.1098/rspa.1921.0029.
4. R. H. Fowler and E. A. Milne, "The intensities of absorption lines in stellar spectra, and the temperatures and pressures in the reversing layers of stars," *Mon. Not. R. Astron. Soc.*, vol. 83, no. 7, pp. 403–424, 1923, doi: 10.1093/mnras/83.7.403.
5. C. H. Payne, *Stellar Atmospheres* (Harvard Observatory Monographs, no. 1). Cambridge, MA, USA: Harvard Observatory, 1925