THE SOLAR CHROMOSPHERE AND THE GENERAL STRUCTURE
OF A STELLAR ATMOSPHERE

Richard N. Thomas

and

Katharine B. Gebbie

Joint Institute for Laboratory Astrophysics*

I. INTRODUCTION

The aim of this symposium is to say thanks to Don Menzel for the ideas and people he has launched on the astronomical world. Useful ideas evolve in a successively more realistic sequence. Productive people ensure such evolution. Whether the resulting picture supports or refutes the original ideas is less important than the generation of such a sequence. Evolving ideas and productive people characterize the Menzel school.

As an example, we outline the evolution of the interpretation of the "solar chromosphere anomaly". This interpretation, which began as a purely solar one, now has broad implications as a key to the general structure of a stellar atmosphere. By "solar chromosphere anomaly", we mean certain observed features, first identified in eclipse spectra, that cannot be reconciled with the classical model of the solar atmosphere. We begin by classifying the essential features of these anomalies and then show how Menzel's interpretation in terms of a collision-free atmospheric appendage has evolved into a more general one in terms of an arbitrary atmosphere diagnosed by what we call the "new spectroscopy". By linking the diagnostic methods of the new spectroscopy to some of Menzel's suggestions on the interpretation of symbiotic spectra, we extend the evolution to a general stellar model. By invoking the concept of transfer and population effects, we distin-

* of the National Bureau of Standards and the University of Colorado.
guish those aspects of the solar chromosphere anomaly that are peculiar to the chromosphere as a distinct region from those that are common to the atmosphere as a whole.

The evolution of the interpretation of the solar chromosphere anomaly reflects a more general evolution in the whole approach to stellar atmospheres. It began with the classical approach, and has, for some of us, culminated in an interpretation of the atmosphere as a transition region between the interior of the star and the interstellar medium. Classically, the atmosphere was viewed simply as the surface of the stellar interior and was assumed to be described by the same state parameters as were used to construct the interior model: pressure, density, and temperature. The single aim of the associated diagnostics was to determine the boundary values of these parameters. Model atmospheres were, by assumption, constructed in terms of these parameters and the relations between them that were used for the interior model. Viewed, however, as a transition region extending from the deepest observable layers to some point at which the star ceases to be the dominant influence on the medium, the atmosphere includes chromospheres, coronas, shells, nebulae, circumstellar dust, and any other phenomena controlled primarily by the star. Here the state parameters and the relations between them must be determined, and a diagnostic procedure developed to obtain the distribution of these parameters from all the available observations. Similarly, the construction of models must be attempted without any a priori assumptions either on the state parameters or the relations between them. In practice, because of our lack of experience with such transition regions, the empirical analysis of observations must go hand in hand with the theoretical construction of models.

Initially the classical atmosphere was thought of as a thin homogeneous surface layer throughout which the interior state parameters were constant at their boundary values. The debate over whether such a model was consistent with the observations first arose over the implications of the existence of a line spectrum. Could a line be produced by scattering in an isothermal atmosphere (Schuster-Schwarzschild) or did its presence require a temperature gradient in the observed layers (Milne-Eddington)? The view prevailed that it did require a temperature gradient, and diagnostic methods based on a non-homogeneous model, but one locally in thermodynamic equilibrium, were introduced to interpret the observations.
The essential characteristic of the classical approach was the use of conventional thermodynamic equilibrium parameters to describe the observable atmosphere. The approach was defined by the five assumptions: local thermodynamic equilibrium for the microscopically distributed functions of matter (LTE-R), hydrostatic equilibrium (HE), radiative or quasi-static convective transport of energy (RE or CE), a spherically symmetric atmosphere (SS), and the absence of macroscopic electric and magnetic fields. The essential physics of this approach is developed in Unsöld's book, *Physics of Stellar Atmospheres*, which, properly supplemented by subsequent developments in atomic physics and computing techniques, remains the classic reference in the field today. The approach has been used to investigate the variation, as a function of Teff, g, and abundances, in the boundary values over the Herzsprung-Russell plane. If the physical basis of the approach were sound, these analyses would provide the essential link between the observations and their interpretation in terms of atmospheric structure and evolution. If, however, the classical assumptions are not valid, the results of these diagnostics provide only a preliminary survey, the meaning of which remains to be assessed.

Not only may the derived values of the parameters be unreliable but in particular their physical significance may also be in question. Such ad hoc parameters as turbulence, excitation and ionization temperatures, and mixing lengths have been introduced to reconcile the observations with the classical model, but the self-consistency and influence of these parameters on the state of the gas has not been satisfactorily explored.

If, on the other hand, the atmosphere is regarded as the transition region between the stellar interior and the rest of the Universe, all those phenomena associated in any way with the domain of influence of the star must be taken into account. The innovation of Menzel and his early collaborators was that in analyzing stellar spectra and constructing stellar models, they implicitly attached as much weight to the chromospheric anomaly, symbiotic spectra, circumstellar shells, and nebulae as to the Fraunhofer spectrum of the star itself. Their early results gave insight into the kind of microscopic interactions that are important in gaseous matter under atmospheric conditions.

At this point, we should perhaps ask whether we are discussing two different but equally valid approaches to a stellar atmosphere, each of which fulfills the purpose for which it was designed. Can the solar atmosphere, for example, be divided into two inde-
pendent regions? The lower region would then produce the observed disk spectrum; its structure would be obtained, and its spectrum analyzed, by the classical approach. The upper region, including all the other phenomena in our more extended definition of an atmosphere, would require more general diagnostic methods and data but would in no way affect either the lower region or the disk spectrum. The significance of the solar chromosphere anomaly lies in the answer it provides to this question.

Until recently, most people would have agreed that such a division of the atmosphere was indeed possible. About a decade ago, however, the results of a long-range program of observation and analysis proved to the contrary, showing conclusively that the chromosphere does seriously affect the disk spectrum (Thomas and Athay 1961). The observational approach was a refinement of that pioneered by Menzel (1931) and developed by his students (Evans, Roberts, and Thomas 1950-51). The diagnostic approach evolved from a viewpoint introduced by Menzel and his associates (Menzel 1931; Cillie and Menzel 1935; Goldberg and Menzel 1948), as extended and refined by his students (Thomas 1948 et seq.; 1950 et seq.). The results of this program showed that (1) as a part of the atmosphere that affects the disk spectrum, we must include a definite region called the "chromosphere", the structure of which is fixed by conditions that violate a number of the assumptions defining the classical model, and (2) there are also parts of the disk spectrum formed in regions below the chromosphere to which the classical diagnostic methods are not applicable.

It is being gradually accepted that neither of these two conclusions is peculiarly solar, but rather that the solar chromosphere anomaly is of great significance for our general picture of a stellar atmosphere. Although the original solar investigations focused on the first conclusion, it is the second that has been more widely accepted in the stellar context. In essence, the second conclusion implies that the assumption of LTE-R should be dropped. It is now becoming fashionable to build models of a classical atmosphere without this assumption; that is, the "New Classical Atmosphere" (Mihalas 1970). That the assumption of LTE-R is untenable was demonstrated by some of us from the Menzel school who, using the solar disk spectrum, showed that LTE-R precluded a self-consistent interpretation first of the chromospheric lines, then of all lines for which the optical thickness at the line center exceeded about one, and then of the chromospheric
continuum. By establishing the theoretical basis for the empirical conclusions, we showed the results to apply to stars as well as to the Sun. We suspect that this conclusion has now become popular mainly because models can in some cases be computed as accurately without LTE-R as with it. The problem remains, however, that again some of us from the Menzel school have shown that the assumptions of hydrostatic equilibrium, radiative (or convective) equilibrium, and spherical symmetry are also invalid in parts of the solar chromosphere, and a great variety of observational work on solar magnetic fields indicates that the remaining assumption of the classical model is also doubtful. Furthermore, we now realize that the whole rocket UV, far-infrared, and the radio spectrum of the Sun all originate in the chromosphere and corona. Thus "chromospheric effects" on the "normal disk spectrum" can be ignored only if the "normal spectrum" is limited to the visible continuum and weak lines. We doubt that even the strongest proponent of the classical atmosphere would maintain such a position.

We conclude then that in any model of the solar atmosphere, some of the classical assumptions must be dropped in almost all of the atmosphere, and all of the classical assumptions must be dropped in some of the atmosphere. We suggest this may also be true for most stellar atmospheres. Is it perhaps significant that model builders are now apparently willing to construct wholly theoretical model atmospheres for any star except the Sun?

In Section II we review the observations and ideas that have led to the rejection of each of the classical assumptions in the solar chromosphere and discuss our replacements for them. In Section III, we suggest how this overall picture of the chromosphere clarifies our general picture of a stellar atmosphere.

II. THE SOLAR CHROMOSPHERE ANOMALY

There are two ways to try to resolve the chromospheric anomaly. The first way is to examine individually the validity of each classical assumption, theoretically and empirically. Unfortunately, we don't know enough about the character of transition regions in astronomical environments to treat the theoretical problem in full generality; and since the assumptions are coupled in their observational consequences, a
wholly empirical approach is ambiguous. A second way to try to resolve the anomaly is to abstract the essential features of all the observed anomalies into a few coherent groups. It is this approach that in practice has proven the more productive. Indeed, the evolution of Menzel's original ideas into our present picture reflects the evolution of the categories into which the anomalies have been grouped and of our understanding of the implications of these categories.

A. The Observed Anomalies

1. Emission Height Gradients. Anomalously low emission height gradients as observed at eclipse in both lines and continua, and the variation in these gradients with excitation and strength of lines, and among the various lines and continua.

Menzel was the first to obtain extensive eclipse data of sufficient precision to allow a quantitative determination of these well-known qualitative effects. Subsequent HAO-Sac Peak eclipse expeditions increased the accuracy and height resolution to the point where differential values of emission height gradients could be investigated as functions of excitation and line strength. At atmospheric densities predicted by the classical model, these observed emission gradients should not be affected by self-absorption and could therefore be identified with the gradients of the occupation numbers. If these gradients were in turn identified with the total density gradient, it would be lower by factors of up to ten than that predicted by a purely thermal classical model; the same density gradient interpreted on the basis of a "turbulence" arbitrarily introduced into the classical model implies supersonic velocities for the turbulent elements. If, on the other hand, the gradient of the occupation numbers is not identified with the density gradient, the implication is that the excitation and ionization increase for some lines and decrease for others in the same atmospheric region.

2. Excitation State. Anomalous relation between emission in lines and continua of the same ion as observed at eclipse, and anomalous intensities and structures of the cores of strong lines as observed on the disk and at the limb.

Eclipse observations in the H- Balmer, and Paschen continua, interpreted simultaneously, give temperatures and densities inconsistent with the classical model. Nor do combined observations of the Balmer lines and continuum give occupation numbers for the
energy states of hydrogen that are consistent with the classical model. Recent observations of the Lyman continuum lead to the same conclusions.

Interpreted on the basis of an unblanketed classical model, disk observations of certain lines (Na I and the hydrogen Balmer lines) imply temperatures of up to about 1300°K below those predicted, while others (Ca II) require that the temperature first decrease, then increase, and then decrease again. The appearance of emission lines on the disk, such as those observed in the rocket UV, has no interpretation on the basis of a classical model.

3. Ionisation State. The observation of ions that should not exist.

The earliest example, often emphasized by Menzel, is the presence of He I and He II lines. Identification of the coronal lines by Edlén intensified the problem, and more recently many more such ions have been identified from the rocket UV spectrum in both the corona and chromosphere.

4. Inhomogeneities. The presence of inhomogeneities in at least the upper chromosphere.

Menzel's early picture of the upper chromosphere was that of a "burning prairie". Later studies at HAO and Sac Peak defined this phenomenon more specifically in terms of the properties of the spicules. Additional evidence is obtained from the anomalous emission in different lines and continua as observed at the solar limb during eclipse.

B. The Implications

This grouping of the observed anomalies suggests the following implications for the structure of the atmosphere and our interpretation of it. (1) Those parts of the atmosphere that contribute to the observed spectrum, including the disk spectrum, extend to greater heights than would be predicted from the solar gravity and effective temperature on the basis of the classical model, which thus represents at most some lower part of the atmosphere. (2) Either the distribution of excitation is quasi-random or this effect is an artifact produced by faulty diagnostics. (3) The energy available for ionization is greater than the classical limit set by the effective temperature. (4) The outer atmosphere is unstable against the production of inhomogeneities. Whether the instability arises in radiative processes or in a differential input of non-radiative energy or momentum or possibly in other sources, remains to be established.
Thus to resolve the chromosphere anomaly, we must identify, describe the operation of, and assess the relative importance of those effects that can cause a real or apparent distribution of absorbing and emitting atoms to differ from that predicted by the classical model. Because the distribution of excited atoms reflects the distribution of both density and excitation, we must specify the factors that fix the density (input of kinetic energy and momentum, and the stability of a homogeneous structure against their differential input) and those that fix the excitation (input of internal energy and its partition). The classical assumptions impose a particular set of relations that fix these distributions in a way that requires no knowledge of the details of the microscopic processes; but in doing so, these assumptions are all coupled in their effect on any empirical test of their validity. Because equipartition is assumed for the kinetic and internal energies, and it is assumed there are no microscopic, non-quasistatic velocity fields, the density and excitation distributions are completely coupled and measured by the distribution of the single parameter, $T_e$. Thus we seek a more general approach in which there is neither coupling between nor restrictions on the distributions of density and excitation. To approach the problem in full generality is impractical. We do, however, have the observational guide that the anomaly increases outward in the atmosphere, whereas the applicability of the classical model improves with increasing depth. Furthermore, there is nothing in the structure of the classical assumptions that incorporates an effect of the boundary. Since the chromosphere, as part of the transition region, appears to have many of the physical properties associated with a true nonadiabatic boundary, a first logical step toward understanding its structure would be to include boundary effects in such a way that the structure produces the chromospheric anomalies at large heights and approaches the classical model at great depths. We now examine the character of a chromosphere viewed as such a transition region.

C. The Chromosphere as a Transition Region

The presence of a non-adiabatic boundary may affect a system in two ways. First, the exchange of energy at the boundary will, in general, result in an energy flux that causes a gradient in the state parameters. Second, the microscopic rate processes that determine the density and excitation distribution func-
tions will become inhomogeneous and so depart from detailed balance.

We consider first the effect of an energy flux. In an adiabatically bounded system, there is no energy flux; the particle concentrations and the distribution of internal energy levels may vary with position, but the temperature and excitation are uniform, fixed wholly by the total energy of the system. In a stellar atmosphere, however, which is characterized by a non-adiabatic boundary and an energy flux, the excitation at each point depends upon: (1) the amount of energy available to be absorbed, (2) the amount of energy actually absorbed, and (3) the process by which the absorbed energy is apportioned over the various energy states. The amount of available energy is fixed by the transport processes, and the amount of energy absorbed is fixed by the density and relative populations of the energy levels, which also describe how the energy is apportioned. Transport and population processes are, in general, coupled. The classical approach restricts the kind of coupling; LTE-R describes the absorption and partitioning of the available energy; radiative equilibrium, hydrostatic equilibrium, and spherical symmetry specify the available energy by eliminating all but radiative transport processes. A more general approach, from which these assumptions are removed, must make possible the calculation of (1) to (3) in terms of the total energy flux and certain other properties of the star that remain to be determined. According to the classical approach, $T_{\text{eff}}$ and $g$ suffice.

We consider next how the presence of the boundary introduces inhomogeneities into microscopic processes causing departures from detailed balance. There are four such effects: (1) Radiative processes reflect the boundary-induced asymmetry in the radiation field and the effects of the resulting discrepancy between color and density of radiation often referred to as dilution effects. (2) Collision rates will decrease with decreasing density; hence there will be an increase in the relative importance of radiation rates. (3) Aerodynamic motions will amplify with decreasing density, ultimately providing a non-radiative energy source with a potential for excitation that is not necessarily related to that of the radiation. (4) These aerodynamic motions, together with any material instabilities associated with radiative effects, may induce geometric inhomogeneities, hence additional boundaries.

To include these boundary effects, we must, in practice, combine two approaches: (1) a synthetic approach, in which theoretical models incorporate these
effects, and (2) an analytical approach, in which a
diagnostics is developed to infer empirically the rela-
tive importance of the effects. In the latter approach,
a major problem has been to avoid imposing diagnostic
methods that, while more general than those of the clas-
sical model, still pre-condition the derived results.
In the following sections, we consider Menzel's origi-
nal suggestions and then show how they have evolved
into our current approach.

1. The original Menzel approach

At the time the Menzel school began to worry about
nonclassical effects, astronomers implicitly assumed
that collisions maintained LTE-R in a stellar atmo-
sphere. Any departure from LTE-R would then be
expected to occur in low-density regions, where, it was
also assumed, the optical depth in all but the Lyman
continuum and lines was sufficiently small that the con-
tinuum radiation field could be described as the prod-
uct of a dilution factor and a Planck function charac-
teristic of the central star. In their initial investi-
gations, therefore, Menzel and his early collaborators
considered only the first and second effects of the
boundary on the microscopic processes in a pure hydro-
gen atmosphere (Menzel 1937 et seq.; Cillie and Menzel
1935). This limitation appeared to be justified by
observation: the excited energy levels in planetary
nebulae and in the solar chromosphere appeared to be
underpopulated relative to the local value of $T_e$, in
accordance with the theory of recombination in a dilute
radiation field.

The electron temperature, $T_e$, in planetary nebulae
was determined by the balance between "dilute" photoion-
izations in the Lyman continuum - a non-LTE process -
and free-free radiative transitions - an LTE process -
in radiative equilibrium. Thus the computed value of
$T_e$ lay below the radiation (or "color") temperature of
the central star, its value in a collision-free region,
but well above its LTE value as determined by the
energy density of the radiation field. Embarrassingly
enough, this approach was not applied to non-nebular
atmospheric regions until 1963 when it was reintroduced
by Cayrel who applied it to the $\text{H}^+$ free-bound continuum
in the solar chromosphere. It is interesting for the
evolution of the classical model that if the collision-
free solution were valid down to $T_e = 1$, and if there
were no LTE processes, the model would closely resemble
that of the Schuster-Schwarzschild isothermal atmosphere.
Such an atmosphere would include everything we have defined as the transition region, including the solar chromosphere and planetary nebulae. The presence of lines would not affect this picture as long as they were formed by pure scattering. Thus we see that deviations from such an isothermal atmosphere arise from collisions, from non-scattering processes in line formation, and from deviations from radiative equilibrium.

In both the prediction and interpretation of the Balmer and Paschen spectra, the error of the Menzel approach lay in assuming the solar atmosphere to be optically thin in the Balmer lines. An empirical analysis (Thomas 1950) allowing for a non-zero opacity in these lines showed the energy levels \( n \geq 3 \) to be over rather than underpopulated. A theoretical analysis (Thomas 1949) showed that the partition of energy among the internal degrees of freedom was controlled by the radiation field in the Balmer lines. Thus the stage was set for the modern non-LTE approach in which the populations of the energy levels and the transfer of radiation are inescapably coupled. Because this approach in terms of a collision-free region with a dilute radiation field suffices for the Balmer, Paschen and higher spectra, it gives information on the partition of energy over the discrete energy levels, but not over the energy levels of the electron continuum. Thus it is relatively insensitive to \( T_e \).

2. Subsequent evolution

The most significant empirical extension of Menzel's work was that of the HAO-Sac Peak eclipse expeditions of 1952 and 1962. The aim of these expeditions was first to repeat Menzel's 1932 observations with greater accuracy and height resolution and second to extend his analysis along the directions just summarized. The primary idea was to adopt the above approach to an empirical analysis of the Balmer and Paschen continua and lines, but to treat the Lyman region theoretically (Pottasch and Thomas 1959) in order to obtain the ionization equilibrium for hydrogen. Combined with an analysis of the \( H^- \) continuum, this approach gave an accurate interpretation of the energy in the continua, i.e. of \( T_e \).

At this point, one had for hydrogen an empirical measure of the partial distribution of an unspecified total amount of energy. By not forcing the value of this total energy to be that of the radiation field, but rather by comparing the empirical \( T_e \) with values
limited by radiative equilibrium, one obtained an empirical measure of the validity of radiative equilibrium. It is interesting that this empirical analysis, incorporating for the first time a correct treatment of coupled line and continuum effects, produced a $T_e(h)$ distribution toward which the various subsequent models seem to be converging (Utrecht Reference Photosphere 1964; Bilderberg Continuum Atmosphere 1968; Revised Bilderberg Continuum Atmosphere 1969; Harvard-Smithsonian Reference Atmosphere 1971). In Figure 1, we reproduce this model (Thomas and Athay 1961) together with other representative models. The outstanding success of this model was that its wholly theoretical prediction of the Lyman continuum of hydrogen came within a factor of two of the observed value. The subsequent models, all of which include some part of the classical assumptions, still failed by several orders of magnitude to predict the Lyman continuum. The current models based explicitly on the Lyman continuum (Noyes and Kalkoven 1970) and one including the effect of spectral lines in the rocket UV (Avrett, Vernazza, and Chipman 1970) predict a $T_e(h)$ distribution very close to that of the 1961 model.

In extending the non-LTE approach to the Lyman region, one had to take into account the third effect of the boundary on the microscopic processes; namely, a possible difference in energy between collisional and radiative processes. In doing so, one had extended the Menzel approach to include all such effects except a geometrical departure from spherical symmetry. And even this effect could be studied from the eclipse observations by comparing the free-bound emissions in hydrogen and He I (Athay and Menzel 1956), and the emission in neutral and ionized metals (House 1961). These comparisons provided evidence for departures from spherical symmetry at heights of about 1500 km and possibly even as low as 750 – 1000 km.

To analyze Balmer and higher lines in the disk spectra, we extended the Menzel approach to opaque but still collision-free atmospheric regions. The evolutionary step was the recognition that it was necessary to solve equations of radiative transfer in order to obtain either the source function or the occupation numbers of the energy levels; that is, it was finally accepted that radiative processes in the line dominated all other terms in the source function. In the transfer equation for the radiation field in the line, the photoionization terms became recognized as source and sink terms rather than as competitive processes in fixing the numerical value of the source function at a
Figure 1. PSC - "Physics of the Solar Chromosphere" (Thomas and Athay 1961); URP - Utrecht Reference Photosphere; BCA - Bilderberg Continuum Atmosphere; HSRA - Harvard-Smithsonian Reference Atmosphere; Linsky and Avrett (1970). Height zero is at $\tau_{5000} = 1$. In placing the PSC model on the same scale, we have shifted its height zero by 410 km.
Figure 2. PSC - "Physics of the Solar Chromosphere" (Thomas and Athay 1961); URP - Utrecht Reference Photosphere; BCA - Bilderberg Continuum Atmosphere; HSRA - Harvard-Smithsonian Reference Atmosphere. Height zero is at $T_{5000} = 1$. In placing the PSC model on the same scale, we have shifted its height zero by 410 km.
Figure 3. PSC - "Physics of the Solar Chromosphere" (Thomas and Athay 1961); URP - Utrecht Reference Photosphere; BCA - Bilderberg Continuum Atmosphere; HSRA - Harvard-Smithsonian Reference Atmosphere. Height zero is at \( \tau_{5000} = 1 \). In placing the PSC model on the same scale, we have shifted its height zero by 410 km.

given point. Because the source function is thus "controlled" by photoionization rather than by the local value of \( T_e \), the population of the Balmer ground state increases and with it the importance of self-absorption effects. Thus both empirical analysis and theoretical computation indicated an over rather than an underpopulation of the energy levels, as had been implied in the earlier investigations. The change came, however, from
carrying Menzel's original approach to its logical conclusion.

In contrast to the above extension of Menzel's collision-free approach, the interpretation of the disk profiles of such lines as $\text{H}$ and $\text{K}$ of Ca$^+$ and Mg$^+$ (Jefferies and Thomas 1959) extended the approach followed in the analysis of the Lyman eclipse data. This represents an application of the third effect of the boundary on the microscopic processes, greater energy in collisional than in radiative excitation, which reflects a coupling between departures from radiative equilibrium and LTE-R. However, we found that, as with the photoionization effect for the Balmer lines, this collisional effect was manifested through the source and sink terms in the transfer equation for the line and not, as in LTE-R, directly in the source function. Thus both collisionally controlled and collision-free treatments of the disk spectrum could be incorporated into the same general framework (Thomas 1957; Jefferies and Thomas 1959 et seq.; Thomas 1965; Jefferies 1968).

In a rough sort of way, this uncouples the departures from the various classical assumptions. The non-LTE approach gives the general form of the source function

$$\text{source function} = \frac{\text{radiation field considered}}{1} + \text{source terms} + \text{sink terms}$$

together with expressions for the source and sink terms as functions of the state parameters. To determine the permitted range in values of these state parameters and hence of the source and sink terms, we study departures from hydrostatic equilibrium, radiative equilibrium, and spherical symmetry.

3. Integration of theoretical investigations

Just as the various observational anomalies could be grouped into four classes, so the different theoretical investigations can be brought together into two categories. Those in the first category, which has been called the "new spectroscopy" (Thomas 1965), aim at classifying the types of source and sink terms that exist and identifying them with the kinds of profiles of lines and continua they produce in different atmospheres. As such, this approach is concerned with what appears to be the unsystematic distribution of excitation with height. The investigations in the second category aim at determining those factors that fix the val-
ues of the state parameters and hence at selecting those components of the source and sink terms that control the source function in any given situation. This approach is concerned with the absolute level of excitation and ionization, the density distribution, and possibly the departure from spherical symmetry. The first approach has been developed to a considerable extent; the second is in a less advanced stage.

(a) The new spectroscopy. This approach arose from attempts to systematize the methods for predicting the kind of profile expected of a given line or continuum in a given atmosphere. It consists of classifying the various transitions, ions, and atmospheres according to the physical processes controlling the source function. Such a classification is essential in identifying those atmospheric parameters about which information can be obtained from the analysis of a given spectral line or continuum.

In LTE, the spectroscopic behavior of a small volume element depends only on the local values of temperature, density, and chemical composition; except for the effect of electron scattering, it does not depend on the radiation field. The observed spectrum depends only on the distribution of temperature throughout the atmosphere. Density and composition determine the atmospheric region, hence the range in $T_e$, that is observed in the line or continuum of a given ion. Thus for a fixed chemical composition but an unspecified distribution of temperature and density, the relation between the location of a volume element and its spectroscopic state is two dimensional. But if to the assumption of LTE, we add the other assumptions of the classical model, we force both the temperature and density to decrease monotonically outward. Then the relation between the position of a volume element and its spectroscopic state is one dimensional. In this case, a sequence of lines and continua arranged according to their depth of formation would show a monotonic change in excitation, and conversely.

Without the assumption of LTE, the spectroscopic behavior of a small volume element depends not only on the local values of the temperature, density, and chemical composition, but also on the radiation field. The radiation field, in turn, depends on the physical properties of the surrounding atmospheric regions, in particular on the distribution of source and sink terms. The kind of source and sink terms that control a given line or continuum will depend on the properties of both the ion and the atmosphere. Hence there is no longer a one-to-one correspondence between the excitation of a
line and its depth of formation, nor do the usual criteria for excitation level necessarily apply. For example, absorption lines with emission cores (collision dominated) may arise in the same atmospheric regions as absorption lines without emission cores (of which some are photoionization dominated). Thus the apparently random behavior of excitation with height - an important feature of the chromospheric anomaly - reflects simply the dependence of the spectroscopic state of a volume element on its environment and on the ionic configuration of the gas. The aim of the new spectroscopy is to develop systematic methods for introducing order where there appeared to be none. We study the apparent excitation level as a function of position in an atmosphere in which the distributions of $T_e$, $n_e$, and chemical composition are specified.

We first determine whether there are atmospheric regions in which collisions dominate over radiative processes in all transitions. If so, these regions are in LTE-R. In most stars, any such regions occur only in the deepest photospheric layers; for some stars, they may occur throughout most of the photosphere; but they never occur in a chromosphere. In those atmospheric regions that are not collision dominated, there are, then, transitions with source functions that depend on the radiation field in the transition. Thus the distribution functions are coupled to the radiation field, and it is necessary to specify the details of this coupling. These details will depend on whether the atmospheric region produces its own radiation field in the given transition, i.e. on whether its optical thickness is greater or less than about one in the transition. If less than one, the source function is controlled by the external radiation field, and we study the region producing this field: for example, in all but the Lyman region, the radiation field in a planetary nebula is produced by the central star. If, however, the optical depth in the transition is greater than one - as occurs often in the solar chromosphere and the cores of all strong lines on the disk - the radiation field is determined by the distribution of source and sink terms. The aim then is to identify these terms and hence the processes, collisional or radiative, that fix them for the particular transition in a given atmosphere.

Since, in general, the ion must be treated as a whole, and since the source and sink terms for its various lines and continua may be coupled, it is often difficult to isolate individual processes. To use a line as a diagnostic tool for the atmosphere, we must obtain either an algebraic expression for the source and sink
terms or a sufficient number of numerical solutions to distinguish the significant processes in fixing these terms. Many "devices" such as the equivalent two-level atom (Thomas and Athay 1961) have been introduced to facilitate the algebraic approach, but whichever approach is used, the formulation in terms of source and sink terms enables us to place collision-controlled source functions, photoionization-controlled source functions, and various mixtures of the two all in the same framework. In the solar atmosphere, the prototypes of these three kinds of source functions are, respectively, the Ca II lines, the Balmer lines, and the Lyman lines. In other atmospheres, the behavior can be quite different. The prototypes have, however, enabled us to check the general structure of the theory and to verify, theoretically and empirically, the influence of the four boundary effects on the microscopic processes.

Thus the new spectroscopy tells us what governs the partition of energy and hence the type of spectrum for a given atmosphere. But to study the detailed variation in the behavior of the lines and continua across the spectral sequence, we need to know the distribution of state parameters in each class of atmosphere. The purpose of analyzing the observations is not to determine the partition of energy in a prescribed atmosphere, but to determine the properties of the atmosphere simultaneously with the partition. The development of methods to do this without preconditioning the result has followed two lines: one empirical, exemplified by the analysis of solar eclipse data; the other theoretical.

(b) An empirical approach to state parameters.
This is an extension of Menzel's approach to a quantitative analysis of the solar chromosphere from eclipse observations of the hydrogen lines and continua. The aim is to obtain an empirical distribution of electron temperature, density (which is essentially the hydrogen concentration), ionization, and excitation, and thus to infer the energy input, momentum input, and partition of energy. In order to do this, we require the distributions of $T_e$, $n_H$, $n_e$, $n_p$, and $b_j$ for all energy levels $j$. Observations in two continua plus lines, a knowledge of the relative abundances of metals to hydrogen, and a theory for $b_1$ and $b_2$ of hydrogen make the problem determinate. In constructing an applicable theory, it is essential that the expressions for $b_1$ and $b_2$ be independent of any assumptions on the input of momentum and energy and that they include all relevant radiation fields, which in practice means the Lyman, Balmer, and
Paschen continua and H α. The chromosphere is transparent in the Balmer and Paschen continua, but the transfer problem in the Balmer lines and in the Lyman continuum must be included. This treatment showed for the first time the effect of this coupled transfer solution on $b_1$ and $b_2$, hence on the empirical determinations of $T_\alpha$. The model obtained from this analysis is included in Figures 1, 2, and 3.

The conclusions of the 1961 analysis most relevant to this paper are as follows: (1) Because the values of $T_\alpha$, $b_1$, $b_2$, and the Lyman continuum radiation are strongly coupled, each depending on an analysis of the atmosphere as a whole, the coupling in the tests for radiative equilibrium and LTE-R stand out. Departures from radiative equilibrium occur as low as 500 km (below this height the temperature distribution is not sufficiently reliable for us to draw definite conclusions). Departures from LTE-R occur in all the observed regions. (2) The physical reliability of the results is confirmed by comparison with independent data. The intensity of the Lyman continuum predicted from the empirical model is within a factor of two of the observed value. He I and He II lines are observed just where the model shows a steep rise in $T_\alpha$ to some value exceeding 10,000°K. (3) The analysis breaks down at the start of this abrupt rise in $T_\alpha$. Empirically, the combined He I and Balmer continua require a geometrically inhomogeneous structure in the region where He I appears. Theoretically, the assumption of detailed balance in the Lyman lines is questionable in the same region. These conclusions on the Lyman continuum have been confirmed recently by Noyes and Kalkofen (1970) who used data in the rocket UV. By treating the Lyman lines and some lines of C II, Avrett, Vernazza, and Chipman (1970) have re-examined and improved the theory in the region of the steep rise in $T_\alpha$.

The Standard Reference Models included in Figures 1, 2, and 3 are an amusing contrast to our empirical model. No account has been taken of eclipse data. The assumption of hydrostatic equilibrium has been applied to what appears to be a freehand extrapolation of $T_\alpha(h)$ in the region $\tau_s \leq 0.05$. Non-LTE effects have been either neglected entirely or included in a patchwork, non-consistent fashion. The original justification for constructing such models was that a standard model would provide a means whereby various theories of line formation could be compared. However, it was perhaps inevitable that once published, these models were accepted as valid representations of the Sun. The construction of numerically accurate but physically incon-
sistent models is in direct contrast to the approach of the Menzel school. It is interesting that these reference models appear to be approaching the 1961 model, which, while it will certainly be modified in detail, seems to be holding up quite well.

(a) A theoretical approach to state parameters: population and transfer effects. The new spectroscopy gives the spectroscopic state of a small element of gas for a given distribution of state parameters. An empirical analysis, such as that described in the previous section, suggests how we might obtain these distributions when we have the range of observations available for the Sun. The question remains of how to obtain them under less favorable conditions. With the exception of certain eclipsing systems, this must be done theoretically.

The values of the state parameters are in general fixed by a combination of transfer processes and local interactions. A solution of the transfer problem throughout the atmosphere provides a value for the transfer quantity at a point, e.g., the energy in the radiation or velocity field; the interaction with the local population of energy levels gives a value for the energy absorbed. Interactions among the energy levels determine the final partition of this absorbed energy. Although these steps are coupled, we can obtain a clearer physical picture by isolating the physical processes and quantities that determine the character of each step. For example, for a grey atmosphere in LTE-R and radiative equilibrium, $J_v$ is determined by the distribution of $T_e$. To a first approximation, $T_e$ is fixed by the local value of the integrated $J_v$. Thus $J_v$ is coupled to the distribution of $T_e$, but neither $T_e$ nor $J_v$ is coupled to the $v$-dependence of $J$.

We attempt to isolate the above steps by separating what we call population and transfer effects. Thus far we have applied our formulation only to the determination of $T_e$. In doing so, we have distinguished between changes in the local value of $T_e$ due to changes in the energy supply, which we call transfer effects, and changes in the value of $T_e$ due to changes in the relative importance of the microscopic processes fixing the populations, which we call population effects. For radiative equilibrium and a one-level atom, we have expressed the kinetic energy of the photo-recombining electrons, which depends on $T_e$ alone, as the product of two factors: (1) the rate at which energy is supplied to the electron continuum per photo-ionized electron, and (2) the number of photoionizations per recombination. The first factor depends on

104
the energy transferred and controls the transfer effects; the second depends on the populations and controls the population effects. It is the behavior of the population factor that determines the explicit dependence of $T_e$ on the process determining the ionization equilibrium. For this reason we have called it the temperature control bracket. For a multilevel atom, we take the weighted average of the photoionized and photorecombined electrons in each continuum. The weighting factor is the product of the temperature control bracket and the relative numbers of photorecombinations or photoionizations in each continuum. Thus far, the applications of this approach have been only exploratory. We believe, however, that it will give the same physical insight into those factors that determine the values of the state parameters as the new spectroscopy gave into identifying types of observed line profiles with classes of source and sink terms.

III. A GENERAL STELLAR MODEL

The interpretation of the solar chromosphere anomaly in terms of the physical processes associated with a transition region suggests that the conclusions of Section II on atmospheric structure and diagnostics are not peculiarly solar but relate more generally to stellar atmospheres. These conclusions may be summarized as follows, corresponding roughly to the classes of observed chromospheric anomalies. (1) The observed disk spectrum is formed in an atmospheric region that extends to far greater heights than that predicted by the classical model, which represents at most the deepest layers of the atmosphere. (2) The major contrast between the spectrum predicted by our method and that by the classical approach is that our method predicts a non-uniform variation of excitation with height in the atmosphere. This corresponds on the one hand to a diversity in the types of source functions, which is associated with differing source and sink terms, and on the other hand to the diversity in non-radiative energy supplies and departures from hydrostatic equilibrium and spherical symmetry. (3) To understand the complete structure of this extended region, it is necessary to allow for departures from all the classical assumptions successively with increasing height. The details can be expressed in terms of population and transfer effects; the former is associated with departures from LTE-R near the boundary, the latter with the production
of a chromosphere. (4) The initiation of any horizontal inhomogeneities by atmospheric instabilities introduces the possibility that any effect associated with a variation in height may also occur in the horizontal direction.

We suggest that just as with the Sun, what appears to be anomalous in stellar spectra is actually the key to the structure of the atmosphere. "Peculiar" spectra are then seen to indicate not something out of the ordinary but simply accentuated features of atmospheric structure common to all stars. The extreme peculiar spectra, known as "symbiotic", represent an enhancement of the same combination of high and low excitation effects already encountered in the solar chromosphere. This suggestion again reflects an early idea of Menzel's, which has been overlooked in recent years. The details of his idea differ from ours, but the basic physical point is the same.

A. Menzel on Symbiotic Stars

Symbiotic stars are those whose spectra include both high and low excitation lines and continua. Many years ago, a number of people, including Menzel, suggested that the obvious interpretation was in terms of binary stars, one hot, one cool. Subsequently Menzel modified this picture by suggesting that such stars were not in fact binaries but single stars divided into two regions: a central, hot nucleus surrounded by an extended, cool atmosphere supported by a combination of radiation pressure, turbulence, and stellar rotation. If the atmosphere were optically thin, the photosphere would be formed close to the hot nucleus, and we might observe a Wolf-Rayet type spectrum. If, on the other hand, the atmosphere were optically thick, the photosphere would be formed near the outer boundary, and the continuum would mimic that of a red giant. He elaborated this model to suggest that a single rotating star could produce both a hot and a cool continuum. The equatorial regions of the atmosphere, distended by rotation, would be optically thick, so the photosphere would be formed in the outer, cooler regions, whereas the polar regions would be optically thin, so one would see the hot continuum characteristic of the stellar nucleus.

We suggest that Menzel was correct in attributing symbiotic spectra to a single star rather than a multiple system. We would, however, alter the details of his model and greatly extend the number of stars to which it applies. Our picture is the following: All
stars are symbiotic, the differences being only a matter of degree, reflecting the relative importance of the classical and non-classical regions of the atmosphere in producing the observed spectrum. The symbiotic properties arise in two ways. In stars without transfer effects, and therefore without chromospheres, they originate in population effects. In this case, the high and low excitation lines cannot be associated with any distinct regions. In stars with transfer effects, and thus with chromospheres, there will be a height asymmetry in ionization, which will, in fact, be the reverse of that suggested by Menzel but in the direction of that suggested by one of us for Wolf-Rayet stars (Thomas 1950) and by Mrs. Payne-Gaposchkin for the Sun (1956 unpublished). You will recall her remark that on the basis of rocket UV spectra alone, she would classify the Sun as a WC6 star, a remark that suggested a similarity in the origin of the two spectra. Thus from disk observations, at least one star, the Sun, appears to be a cool star as seen in the visual spectrum, which originates in the lower atmospheric regions, and a hot star as seen in the UV, which originates in the upper regions. The principle is the same as Menzel's: symbiosis reflecting the schizophrenia of a single star. The reversal of the hot and cold regions comes from experience with Menzel's favorite star, the Sun, and follows the direction that he himself took in exploring the He I and He II anomaly. He can hardly object to this change in detail.

B. A General Stellar Atmosphere

In the deepest layers of the atmosphere, we have a region represented by the classical model. Here LTE-R is established by the predominance of collisions over every other kind of excitation process. The source function is fixed by $T_e$, which is in turn fixed by either radiative or quasi-static convective energy transport. As we move higher in the atmosphere, collisional processes decrease relative to radiative, and eventually we reach a point where radiative processes dominate. Here we may still have LTE-R in limited spectral regions provided the important radiative processes maintain a homogeneous radiation field. Effectively this means that no radiation can be observed in these transitions, but other transitions, to which these are coupled, can be observed. On this basis, we impose conditions on the maximum value of $T_e$ and the minimum value of density, hence of gravity and momentum supply,
for the existence of an observable LTE-R region (Gebbie and Thomas 1971). It cannot be assumed a priori, how-
however, that this region represented by the classical
model can in fact be observed.

Above the region represented by the classical
model, we must allow for departures from any or all the
classical assumptions. Whether departures from LTE-R
occur can be computed in terms of density and optical
depth, if we know the distribution of state parameters.
Thus, as we have stressed, we must investigate both pop-
ulation and transfer effects, either empirically or
theoretically or both. In one atmosphere, we may have
only population effects; that is, we may have depart-
tures from LTE while the other classical assumptions
remain valid. In another atmosphere, with transfer
effects, several of the assumptions may break down
simultaneously.

If in the region above that described by the clas-
sical model, there are only population effects, we call
it a non-classical photosphere. Here $T_e$ may rise to
some maximum value limited by the value of $T_{eff}$. The
possible presence of several kinds of source and sink
terms may give rise to the apparent anomalies in excita-
tion reflected in the disk spectra.

If, on the other hand, this region has both popula-
tion and transfer effects, we call it a chromosphere.
(We do not distinguish between a chromosphere and coro-
na.) Here $T_e$ will rise to a value limited only by the
input of non-radiative energy. Our understanding of
this process is not yet sufficient to predict this maxi-
mum value of $T_e$ in terms of the observed characte-
ristics of a star. Thus the chromosphere is characterized
by the presence of transfer effects, of $dT_e/dh > 0$, and
of $T_e > T_{eff}$.

We now consider how we might establish at least
the existence of a chromosphere from the observations.
From the continuum formed in the deepest atmospheric
levels, we can try to fix the general excitation level
of the spectrum. This corresponds roughly to the effec-
tive temperature of the classical model. We then try
to estimate the variation of excitation level with
height using as our criteria the diversity of the spec-
tra. If this excitation level exceeds that of the
limit set by population effects alone, we conclude that
transfer effects are contributing and that we have a
stellar chromosphere.

For example, from the visual continuum of the Sun,
we obtain an excitation level of about 5800°K for the
regions represented by the classical model. Moving out-
ward through the regions in LTE-R, $T_e$ decreases to some-
thing less than 4700°K - current estimates favor a value of about 4300°K. In the non-LTE regions, population effects could raise $T_e$ to 5300 – 5500°K in radiative equilibrium (Gebbie and Thomas 1970). The observation of He I in the disk spectrum by itself suggests an excitation level close to 10,000°K. The central emission cores of Ca+ H and K tend to confirm that the excitation level rises above the limit set by population effects alone. The rocket UV and radio spectra also support this conclusion. Thus from disk spectra alone, we infer the presence of transfer effects and hence of a chromosphere. Until we obtain a satisfactory treatment of the aerodynamic problem, we must, of course, determine the final model empirically.

Another example is the Wolf-Rayet stars. From the visual continuum alone, we infer an excitation level of some 40,000°K, give or take 10,000°K. Yet from the line spectrum in the visual region, we infer excitation conditions corresponding to some 200,000°K. Again we infer the presence of transfer effects and a chromosphere.

These examples can be extended to other stars as well as various stages of novae and binary systems consisting of an extended supergiant and a small hot companion. We believe this provides a picture of how Menzel's original ideas on the solar chromosphere, flavored by suggestions on symbiotic spectra and passed through the strainer of non-LTE diagnostics, have led to suggestions for a general model for a stellar atmosphere.

ACKNOWLEDGMENT

We are indebted to J-C. Pecker for continuous interchanges on the ideas in this paper and to Alice Levine for valuable advice on the expression of them.

BIBLIOGRAPHY


Bilderberg Continuum Atmosphere, Revised, Gingerich O. 1969, as reported at AAS Solar Physics Meeting, Pasadena.


Thomas, R.N. 1965, Non-Equilibrium Thermodynamics in
the Presence of a Radiation Field (Boulder: University of Colorado Press).
Utrecht Reference Photosphere, Heintze, J.R.W., Hubenet,