Astronomical Data Analysis Software and Systems VIII
ASP Conference Series, Vol. 172, 1999
D. M. Mehringer, R. L. Plante, and D. A. Roberts, eds.

Systemization and Use of Atomic Data for Astrophysical Modeling

Nancy S. Brickhouse

Smithsonian Astrophysical Observatory, Cambridge, MA 02138

Abstract. The growth of supercomputing capabilities over the past decade has led to tremendous opportunities to produce theoretical atomic data for use in modeling astrophysical plasmas; however, not all data are created equally. Both accurate and complete atomic data are important to astrophysical modeling via plasma codes. Critical theoretical evaluations, experimental benchmarks, and even astrophysical observations themselves are all useful in assessing the temperatures, densities, opacities, abundances, and other physical quantities determined from spectroscopy. The database challenge for the future is to move toward increasing both flexibility and standardization of atomic data formats so that the quality of the data can be taken into account in modeling. Problems in X-ray spectroscopy will illustrate the importance of opening up what has often been a black box to many astronomers.

1. Introduction

Spectroscopists in astrophysics may be heard to remark, "If one picture is worth a thousand words, then one spectrum must be worth a thousand pictures." While ultimately the goal of spectroscopy is to determine physical properties and discern among physical processes, the levels of inference involved in "diagnosing" the plasma tend to make the art of spectroscopy less than easily accessible to the general astronomy community. A spectrum's value lies in the quality of the underlying models.

The interpretation of spectra is built upon models of the astrophysical source (e.g., the magnetic confinement of hot plasma in stellar coronae), of the plasma processes (e.g., the equilibration of electron velocities through collisions to achieve a Maxwellian distribution), and of the atomic physics (e.g., the representation of the microphysics of atomic rates). Thus, determining the average density of an emitting source from a measured diagnostic line ratio may appear to be straightforward, but the reliability of such a density is difficult to assess without accounting for all the levels of modeling. Furthermore, the reliability of a diagnostic also depends on its observational context, not only signal-to-noise and calibration, but also the degree of contamination by line blending. The isolation of a diagnostic feature from other spectral components thus becomes an additional modeling concern.

The application of astronomy to fundamental spectroscopic studies has its origins with the Fraunhofer lines in the Sun in the early 1800's, about the same time as the beginning of quantitative laboratory work. Basic spectroscopy continues to be important to astronomy as we probe the universe with increasing spectral coverage, spectral resolution, throughput, and calibration. Atomic data that result from fundamental theoretical and experimental work then become critical components for astrophysical modeling.

Calculation, compilation, and distribution of atomic data are increasingly becoming the responsibility of the astrophysics community, as our needs diverge from the primary interests and activities of atomic and laboratory plasma physicists. Systemization will enhance the value of atomic data for astrophysical modeling. While each field of astrophysical spectroscopy has a diverse set of needs and expectations for spectral analysis, we expect that some general design principles are likely to apply. This paper discusses some of these design issues, using our modeling efforts in high energy astrophysics as an example. The ongoing development project for the Astrophysical Plasma Emission Code (APEC) is implementing these ideas (Smith et al. 1998).

2. Approaches to Spectral Analysis (with a Bias toward High Energy Spectroscopy)

Two complementary approaches are used routinely to analyze spectra. One approach is to choose one or two lines from which to derive a physical quantity, such as density or temperature. This approach has the advantage that bright isolated lines with good theoretical models may be selected. On the other hand, this approach can lead to errors if the plasma is not isothermal or isochoric. Furthermore, the theoretical models may not be as reliable as one might hope. Finally, to test for consistency among derived quantities, one needs to consider other information, often in the same spectrum.

A second approach is to construct a model with sufficient numbers of variable parameters that a good fit to the entire spectrum can be found. This "global fitting" approach is still widely used in X-ray astronomy because only moderate resolution ($\lambda/\Delta\lambda \sim 50$) has been obtained beyond the Sun. Even the strongest emission lines are significantly blended with continuum and other lines at this resolution, as shown in Fig. 1. Although this approach utilizes the entire spectrum and is easily automated, disadvantages are inherent. The information content of the spectrum is not necessarily reflected in bin-based fitting since the model parameters are not independent. Furthermore, methods of this type place a huge burden on the accuracy and completeness of the atomic data needed.

X-ray spectra which are to be obtained with the *Chandra X-ray Observatory* gratings will cover a broad band, providing enormous information content in each observation. An integration of both approaches is needed: to derive initial models from the strong, well understood lines and to test these models against the entire spectrum for consistency or problems. To make the atomic data and spectral emission models available for both purposes, we need to open up the black box of spectral analysis.

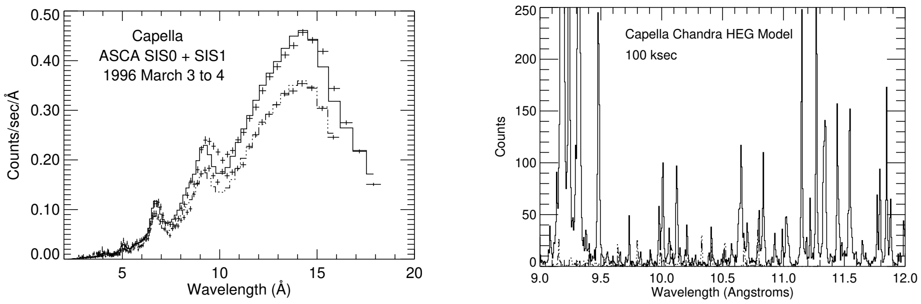

Figure 1. Global fitting methods are used to analyze moderate resolution spectral data, placing high demands on the accuracy and completeness of the plasma emission model. *Left*: Models of the ASCA spectrum of Capella show a deficit of flux around 10 Å (Brickhouse et al. 1999). New calculations of Liedahl & Brickhouse (1999) show that the flux deficit is removed if additional lines from high principal quantum number $n > 5$ of Fe XVII to XIX are added to the model. *Right*: Simulations of the *Chandra* High Energy Transmission Grating observation show a wealth of emission lines blended at CCD resolution. The new lines of Liedahl & Brickhouse are shown separately as dash-dotted lines.

3. Opening Up the Black Box

The spectral codes developed in the 1970's for hot, collisionally ionized plasmas compute the fundamental rates of ionization/recombination and excitation/decay processes (Raymond & Smith 1977; Raymond 1998; Mewe 1991 and references therein). While the original codes model most of the important processes, increased computing capacity over the past two decades has led to improvement in the accuracy of many atomic rate coefficients and to an increase in the numbers and detail of available rates, as illustrated in Fig. 2.

Modular design allows the atomic data, formerly buried in Fortran DATA statements, now to be stored in a stand-alone database. Accommodating simple changes to atomic data should be much easier than in the past. The increase in functionality, discussed in more detail below, is enormous.

We are currently compiling the Astrophysical Plasma Emission Database (APED) for high energy spectroscopy as a set of FITS files. These files are easily viewed, accessed, and modified with standard software. Multiple sources of data are supported, as are multiple functional forms for data. APED will facilitate identification of spectral lines in complex spectra, comparison of atomic data, access to reference material associated with the data, and other spectral analysis functions.

3.1. Spectral Synthesis Codes

Spectral synthesis codes calculate the relevant atomic processes leading to line and continuum emission, under assumptions about the physical conditions. Examples include codes to model stellar atmospheres, photoionized plasmas, and

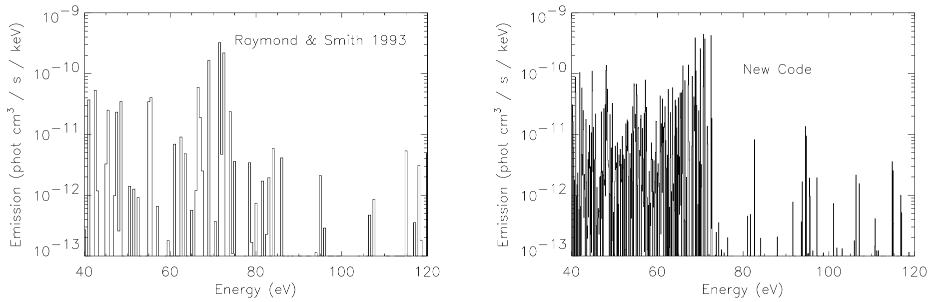

Figure 2. The number of emission lines currently included in new models greatly exceeds the numbers from previous work. *Left*: A Raymond-Smith model using the 1983 version of the code. *Right*: APEC model (Smith et al. 1998), using atomic data compiled by the Chianti collaboration (Dere et al. 1998) for the same spectral band. Not only does the new model show more emission lines, but strong features do not coincide because the older model uses multiplet level averages rather than individual fine-structure levels.

collisionally ionized plasmas. The predicted spectra are compared to observations using a set of physical parameters, which are then adjusted to obtain the best fit. The relevant observations can either be a select set of diagnostics or the entire spectrum (and even beyond). The physical parameters obtained in this way information may be temperature, density, ionization state, structure, abundances or other parameters constructed from these characteristics.

Typically the model parameters may be both underdetermined and overdetermined by the observed spectrum. Using the example of a collisionally ionized plasma, let us take the simplest case of a "coronal model." The coronal model assumes that the plasma is in ionization equilibrium, is optically thin, and is of relatively low density. If we also assume that the observed emission comes from a single temperature, then the "best fit" temperature and its emission measure (i.e., the amount of emitting material) are easily determined from the observed photon flux. Stellar coronal plasmas are not generally isothermal, however, and thus an emission measure distribution (emission measure as a function of temperature) is used to parameterize the coronal structure. Emission lines may come from a range of temperatures in the model, leading to an overdetermined model, or interesting temperatures may simply not be represented by features in the spectral range under observation. Moreover, observed features may simply not be represented by the plasma model. For example, the He II UV and EUV lines observed in stellar transition regions are almost certainly optically thick, and possibly populated by photoionization followed by recombination cascades, rather than by the collisional ionization and excitation processes which dominate most of a coronal spectrum.

The challenges of using a spectral synthesis code to obtain physical information about the source are not only to try to get as much detail as possible correct, but also to understand the limitations of the model. These limitations are generally lumped into the concept of "systematic" or model uncertainties and can be difficult to quantify. Such uncertainties include errors associated with the input atomic rates. For example, collisional data are almost entirely compiled from theoretical sources, so that one must consider the type of theoretical calculation to estimate the uncertainty. The model may be incomplete, by not including all the relevant excitation processes (leaving out a recombination cascade that contributes 20% of the level population), or in its treatment of the model ion (not enough energy levels or lines). Of course, one must also consider the possibility that the model does not apply at all to a particular feature, as in the example of He II above. Finally, of course, there are observational uncertainties as well, which can be either statistical (e.g., counts) or systematic (e.g., calibration; line blending) in nature.

The scientist judges that a model is acceptable if it does a pretty good job of fitting the spectrum. Unfortunately, quantifying what "pretty good" means is only mathematically straightforward in the statistical sense (e.g., by using χ^2). One builds confidence by getting familiar spectral features correct, while one worries that problem areas may either be astrophysically interesting (meaning that the model might not be entirely correct and that some exciting new phenomena must be invoked) or boring (just another inaccurate atomic rate). If one wants to rely on only a few diagnostic lines or line ratios, the problem of distinguishing between oversimplification in the treatment of the emissivity calculation, inaccuracies in the atomic rates, complexity at the source (such as multiple densities when trying to use a density diagnostic), must all be taken into account before the diagnostic is judged reliable. Usually this involves finding consistency with a large set of other diagnostics. Improving scientific judgment on the quality of the spectral fit requires opening up the atomic data and plasma emission models to scrutiny.

3.2. Atomic Database

Improvements in the atomic data that are input to a spectral synthesis code include improvements in both accuracy and completeness. Atomic rate uncertainties range from a few percent to factors of 2 or even larger, depending on the theoretical atomic model approximations used. Quantities such as wavelengths and atomic transition probabilities (or oscillator strengths) are available electronically for many transitions, although lack of standardization makes accessing the information time consuming. Publicly available collisional data tend to be sparse and poorly organized. Standardization might be yet another improvement, but the astrophysics community will need to negotiate standardization with the producers of atomic data and other users.

Converting collisional data into a standard format is a daunting task. As an example, electron impact (or "collisional") excitation data exist in the literature as cross sections, rate coefficients, collision strengths, and Maxwellian-integrated "effective" collision strengths. These quantities are not simple multiplicative factors of each other but can require extensive numerical computation to convert from one form to another. The situation is made still more complex because the data may be stored as a grid of data values or as fits to physically-based functions. Older data do not cover a sufficient energy grid, and thus need to be extrapolated with physical models. The most accurate collisional rate data include the calculation of resonance structure near threshold, requiring cross sec-

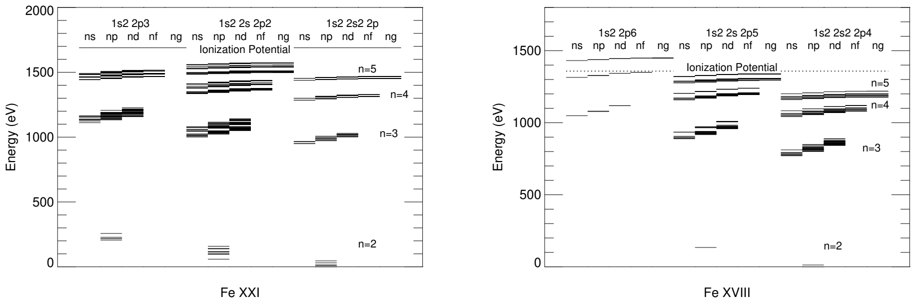

Figure 3. Energy level diagrams from APED illustrate the issue of database size. *Left*: The Fe XVIII model includes 501 energy levels up through $n = 5$ and $\sim$ 13,000 transitions (sparse relative to a possible $501 \times 501 = 250,000$), compared with Raymond-Smith, which lists 11 multiplet lines. *Right*: The Fe XXI model includes 848 energy levels up through $n = 5$ and $\sim$ 16,000 emission lines, compared with 12 lines in a Raymond-Smith model.

tions to be calculated at thousands of energies. These data are then integrated over the electron energy distribution (usually taken to be Maxwellian) to produce effective collision strengths. Current computing power only allows a small number of transitions to be calculated with resonances. Large numbers of lines are obtained by building bigger, but less accurate, atomic models. By combining these two types of theoretical data, we can incorporate the best information for a given ion. With large numbers and diversity of collisional data sets today, an approach which stores data in as original a form as possible, along with keys to the rate calculation method is less labor-intensive and more reliable.

The APED project is compiling data now for $\sim 10^6$ lines of use for highly ionized plasmas, as illustrated in Fig. 3. Even with this number of lines, we now know that we are missing important bands of radiation (e.g., $n > 5$), which, when theoretically calculated and incorporated into the database, will almost certainly be multiplet or quantum-number averaged, rather than fine-structure lines. While the number of lines included now in stellar atmosphere models far exceeds this (e.g., Hauschildt et al. 1997), APED will probably of comparable size. Oscillator strengths constituting the atomic data needed for opacity calculations are constants intrinsic to the spectral line, whereas collisionally-dominated spectral line emissivities vary with temperature and density.

The critical evaluation of atomic data remains an important area of research. Laboratory measurements provide key inputs to the assessment; however, the large numbers of data needed prohibit the exclusive use of experiments to test collisional cross section theory. In X-ray spectroscopy, astrophysical observations are compared with models to determine problem areas, and set priorities for laboratory work (Fig. 3). The *Chandra* X-ray Observatory Center (CXC) is sponsoring a collaborative effort known as the Emission Line Project to produce X-ray catalogs of three bright stellar coronal sources. These catalogs will provide guidance as to the overall quality of the coronal models, identify problem spectral features, and help to set priorities for further theoretical and experimental work.

4. Improvements in Functionality

With an atomic database (including error estimates) that stands alone, the functionality of the spectral synthesis code can be greatly increased. The most accurate theoretical data can be selectively used for line and line ratio diagnostics. Whereas high accuracy for an individual line might be obtained by patching together disparate sources of atomic data, there may be an advantage to using a complete set of atomic data calculated self-consistently (even with lower accuracy) for a global fit model.

For radiative cooling calculations called by hydrodynamics codes, or in non-equilibrium ionization calculations for which the number of unknown parameters is large, more efficient computations are required. Rather than modifying code, one can manipulate the atomic database, producing a smaller averaged database in the same format. Bundling fine structure energy levels following standard atomic physics "sum rules" is straightforward. The accuracy of new models may then be benchmarked against models using the original data.

By tracking accuracy estimates in the atomic database, uncertainty analysis can be folded into the modeling at three levels: by comparing different data sets in the database, by sensitivity testing (varying database values "by hand" to see how sensitive the line emission is), and by Monte Carlo modeling of the entire set. While it is difficult to assess systematic errors in the data (such as might occur, for example, if resonance contributions have been left out), at least data analyzers can be aware of the pitfalls.

Finally, the atomic database as well as models derived from spectral synthesis codes can be used directly in data analysis as table models. The atomic database alone is useful for line identification, for understanding line formation through access to an energy level diagram, and for choosing diagnostics based on accuracy.

5. Conclusions

Not all atomic data are created equally; fortunately, modern storage and computational methods allow us to account for this diversity. Both accuracy and completeness are necessary for spectral modeling. Database formats which allow for different types and qualities of data will improve the integration of data into a critically evaluated set, while flexibility in processing allows for intercomparison of different data within the context of the plasma models.

Since spectral synthesis codes are called on to perform a wide variety of tasks, ranging from calculating diagnostic line ratios to calculating radiative cooling rates needed by hydrodynamics codes, a flexible database can be tailored to the needs of the user.

A range of tools makes possible better assessment of the reliability of astrophysical interpretations, accounting for atomic data, atomic processes, plasma processes, and the quality of observations. Collaboration among people involved in rates, codes, and astrophysics is the best way to address needs in these areas.

Acknowledgments. My colleagues on the development of the Astrophysical Plasma Emission Code (APEC) have contributed enormously to the improvements discussed here. I gratefully acknowledge John Raymond, Duane Liedahl, and Randall Smith for sharing their knowledge, insights, and opinions with me on all levels of modeling. I also acknowledge the support of the Chandra X-ray Observatory Center (formerly the AXAF Science Center) and NASA Grant NAG5-3559.

References

Brickhouse, N. S., Dupree, A. K., Edgar, R. J., Liedahl, D. A., Drake, S. A., White, N. E., & Singh, K. P. 1999, in preparation

Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., & Young, P. R. 1998, in ASP Conf. Ser., Vol. 143, The Scientific Impact of the Goddard High Resolution Spectrograph, ed. J. C. Brandt, T. B. Ake, & C. C. Petersen (San Francisco: ASP), 390

Hauschildt, P. H., Shore, S. N., Schwarz, G. J., Baron, E., Starrfield, S., & Allard, F. 1997, ApJ, 490, 803

Liedahl, D. A. & Brickhouse, N. S. 1999, in preparation

Mewe, R. 1991, Astron. Astrophys. Rev., 3, 127

Raymond, J. C. 1988, in Hot Thin Plasmas in Astrophysics, ed. R. Pallavicini (Dordrecht: Kluwer), 3

Raymond, J. C., & Smith, B. W. 1977, ApJS, 35, 419

Smith, R., Brickhouse, N., Raymond, J., & Liedahl, D. 1998, "High Resolution X-Ray Spectroscopy: Issues for Creating an Atomic Database for a Collisional Plasma Spectral Emission Code", European Space Research and Technology Centre, `http://astro.estec.esa.nl/XMM/news/ws1/procs/smithrk.ps.gz`