BAYESIAN RECONSTRUCTION OF ELECTRON ENERGY DISTRIBUTIONS FROM EMISSION LINE INTENSITIES

R. FISCHER, W. JACOB, W. VON DER LINDEN AND V. DOSE
Max-Planck-Institut für Plasmaphysik, EURATOM Association
POB 1533, D-85740 Garching, Germany[†]

Abstract. Low-pressure plasmas are nowadays widely used for technical applications of plasma-surface interactions, such as plasma etching, material deposition, sputtering, etc. For a thorough understanding of individual processes in plasma processing the electron energy distribution (EED) function in the bulk plasma is of great importance. The EED determines the rates of all electron induced reactions as ionization, excitation or dissociation of molecules. The ubiquitous assumption of a Maxwellian EED becomes progressively worse for hot and low-density plasmas. Measurements of the EED with probes penetrating the plasma result in deteriorating effects on the plasma and the probe, thus measurements without plasma contact are of great interest. A non-destructive measurement is the detection of radiation emitted by the plasma.

The form-free reconstruction of the EED from a small number of measured emission intensities results in an ill-posed inversion problem. In order to avoid spurious features due to overfitting of the data (*ringing*) we apply Bayesian probability theory along with the adaptive-kernel method. The Bayesian approach will be applied to emission lines of helium, since in this case the relevant atomic input quantities are best known.

Key words: Electron Energy Distribution, Low-Pressure Plasma, Inverse Problem, Over-Fitting, Adaptive Kernels, Occam's Razor

[†]Email: Rainer.Fischer@ipp.mpg.de

1. Introduction

The knowledge of the electron energy distribution (EED) in plasmas is essential for modeling the chemical reactions in plasmas, growth processes in thin film deposition, surface treatment with plasmas and wall erosion and material migration in fusion experiments. The relevant quantity is the rate of electron-impact induced processes. This rate is determined by the integral

$$< \sigma v >= \int_0^\infty dv\ f(v)\, v\, \sigma(v) \tag{1}$$

with the EED $f(v)$, the electron velocity v and the cross section $\sigma(v)$. Fig. 1

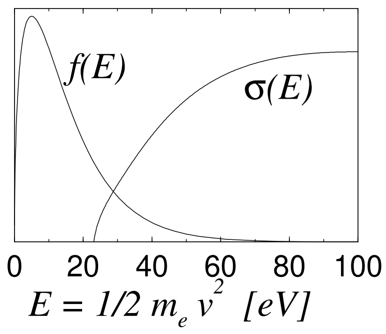

Figure 1. Sketch of a typical electron energy distribution and a typical electron-impact cross-section. The overlap of the two curves overlap only in a small energy range.

shows a sketch of a Maxwellian EED f and a cross-section σ for electron-impact excitation of helium from the ground state. The two curves overlap only in the high-energy tail of the EED. Below the threshold energy of this cross section the rate contains no information about the EED. Excitations between excited states with lower threshold energies couple the EED at lower energies to the rate but this is expected to be rather weak since excitations between excited states are less frequent.

It is common practice to assume thermodynamic equilibrium and hence a Maxwell distribution for the EED. In this case there is only one free parameter, the electron temperature T. The information content in the emission experiment is sufficient to determine T unambiguously. For non-equilibrium situations the generalized Druyvesteyn distribution characterized by the electron temperature T and an exponent ν, $f(E) \propto \sqrt{E} \exp(-(E/kT)^\nu)$, is used [1]. The exponent allows for a steep decrease in an energy region where energy losses due to excitation of atoms or molecules provide a loss channel for the EED. In order to describe multiple energy loss channels a superposition of two or more such functions may be required. In this work the rules of probability theory are used for a form-free reconstruction of the EED.

2. Experiment

The low-pressure plasma is generated by electron cyclotron resonance heating with a microwave. The microwave accelerates those electrons which gyrate in a magnetic field with the frequency of the electric field. The heated electrons leave the resonance zone by drift or impact with other particles. Details can be found in [2].

Two established methods for measuring the EED are given by Langmuir probes and by the detection of radiation emitted by the plasma. A Langmuir probe is a thin wire which penetrates the plasma. The values of the electron temperature and electron density is fitted to the current-voltage characteristic of the wire. An advantage is that the measurement can be done for different locations in the plasma. Disadvantages arise due to the direct contact of the wire with the plasma. Local disturbances of the plasma by the wire can occur and film growth on the probe surface can distort the current-voltage characteristic. The thermal load on the probe restricts the parameter range and interpretation problems in strong magnetic fields can arise.

A measurement method, which is non-destructive for both the plasma and the probe, is the detection of radiation emitted by the plasma. The intensity of an emission line depends on the density of the excited state n_i which emits the radiation integrated over the line of sight,

$$I_{ij} = \frac{1}{4\pi}\frac{hc}{\lambda_{ij}} A_{ij} \int_0^L dx\, n_i \quad , \tag{2}$$

where λ_{ij} is the wavelength and A_{ij} is the transition probability from level i to level j. In this work we have only one line of sight. Thus we have no spatial resolution.

We analyzed the intensity of 8 helium emission lines from principal quantum numbers 3 and 4 to principal quantum number 2 in the photon energy range 2.1-3.2 eV. There are four measured optical transitions in the singlet and triplet term systems of helium, each. The majority of transitions are not recorded due to experimental restriction in the photon energy range.

3. Collisional-radiative model

For determining the intensities of spectral lines, the densities of excited levels n_i are calculated with the collisional-radiative model. $\boldsymbol{n}$ in the stationary state is given by a balance of population and de-population processes:

- e^--impact excitation and de-excitation $i \to j$: $\pm n_i n_e < \sigma v >_{ij}$
- e^--impact ionization $i \to \infty$: $n_i n_e < \sigma v >_{i\infty}$
- population and de-population by radiative decay: $\pm n_i A_{ij}$
- excitation by re-absorption of radiation: $A_{ij} \to \Theta_{ij} A_{ij}$
- de-population by wall collisions: $n_i \Gamma_i$
- dielectronic and radiative recombination: $\beta_{DE/Rad}\, n_{ion} n_e$

n_e denotes the electron density, $0 < \Theta_{ij} < 1$ is the escape factor of radiation from the plasma, Γ_i is the wall-collision frequency and $\beta_{DE/Rad}$ denote the dielectronic and radiative rates. The atomic input data are fairly well known for helium. For a given EED $f(E)$ we have to calculate the rate coefficients $< \sigma v >_{ij}$ of the electron-impact processes and solve the linear equation for the density of excited levels,

$$\begin{aligned} \mathbf{C}\boldsymbol{n} &= \boldsymbol{r} && (3)\\ \mathbf{C} &= \mathbf{C}\left(< \sigma v >_{ij}, n_e, A_{ij}, \Theta_{ij}, \Gamma_i, \beta_{DE/Rad}\right) && (4)\\ \boldsymbol{r} &= \boldsymbol{r}\left(< \sigma v >_{ij}, n_e, n_{He}, \beta_{DE/Rad}\right) && (5) \end{aligned}$$

for the helium density n_{He}. With the assumption of a density profile independent of the level number we can calculate the intensity of the emission lines from eqn. 2. Details of this model can be found in [3].

Since we have only 8 measured data and a form-free reconstruction with about 100 degrees-of-freedom (DOF) we are dealing with an ill-posed inversion problem.

4. Bayesian Probability Theory

The ill-posed inversion problem is tackled by Bayesian Probability Theory (BPT).

In our case of independent measurement errors σ_i, which are estimated from N_d data d_i, a Gaussian likelihood with the χ^2 misfit applies.

$$\begin{aligned} P(\boldsymbol{d}|\boldsymbol{f}, \boldsymbol{\sigma}, I) &= \frac{1}{\prod_{i=1}^{N_d} \sqrt{2\pi}\sigma_i}\, exp\left(-\frac{1}{2}\chi^2\right) && (6)\\ \chi^2 &= \sum_{i=1}^{N_d} \left(\frac{d_i - I_i(f(E))}{\sigma_i}\right)^2 && (7) \end{aligned}$$

In order to apply Bayes' theorem we need a prior probability for the EED f. The appropriate prior for a positive and additive distribution is the entropic prior [4].

$$\begin{aligned} P(\boldsymbol{f}|\alpha, I) &= \frac{1}{Z} exp\,(\alpha S) && (8)\\ S &= \sum_{j=1}^{N} f_j - m_j - f_j\, ln\left(\frac{f_j}{m_j}\right) && (9)\\ Z &= \int d^N \boldsymbol{f}\, P(\boldsymbol{f}|\alpha, I) && (10)\\ P(\boldsymbol{f}|I) &= \int d\alpha\, P(\boldsymbol{f}|\alpha, I)\, P(\alpha) && (11) \end{aligned}$$

$\boldsymbol{m}$ denotes the default model. Marginalization over the hyper-parameter α completes the prior.

It has been demonstrated by the authors [5] that even the optimal treatment of the α-marginalization leaves some residual *ringing* and noise fitting. For form-free density estimation, Quantified Maximum Entropy (QME) yields disappointing results [6]. The reason in both cases is the large number of DOF of a form-free distribution which is not sufficiently restricted by the entropic prior since it contains no correlation between the image cells. We need an approach that reproduces significant structures in the data while noise-fitting by redundant DOF is avoided. In addition, in real world we always have functions containing a mixture of sharp and smooth structures, for example signal and background. We therefore need a method that adopts to the local information content. An Occam factor arising self-consistently from BPT penalizes the complexity of the model reducing the DOF to the minimum amount necessary to describe the data.

Closely related to the preblur concept of John Skilling [4] we convolve a hidden image with a smoothing kernel but with locally varying kernel widths [7].

$$f(x) = \int dy B\left(\frac{x-y}{b(y)}\right) h(y) \tag{12}$$

$$B\left(\frac{x-y}{b(y)}\right) = \frac{1}{\sqrt{2\pi}b(y)} \exp\left[-\frac{1}{2}\left(\frac{x-y}{b(y)}\right)^2\right] \tag{13}$$

We assume a Gaussian kernel. The rules of BPT are used to determine the posterior probability density. To this end we marginalize the hyper-parameters: hidden image $\boldsymbol{h}$, blurring widths $\boldsymbol{b}$ and α.

$$P(\boldsymbol{f}|\boldsymbol{d}, I) = \int_0^\infty \cdots \int_0^\infty d^N\boldsymbol{h}\; d^N\boldsymbol{b}\; d\alpha\; P(\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{b}, \alpha|\boldsymbol{d}, I) \tag{14}$$

$$\propto \int_0^\infty \cdots \int_0^\infty d^N\boldsymbol{h}\, d^N\boldsymbol{b}\, d\alpha\, P(\boldsymbol{d}|\boldsymbol{h}, \boldsymbol{b}, I)\, P(\boldsymbol{f}|\boldsymbol{h}, \boldsymbol{b}, I)\, \times \tag{15}$$

$$P(\boldsymbol{h}|\alpha, I)\, P(\boldsymbol{b}|I)\, P(\alpha|I) \tag{16}$$

This can be done by Markov-Chain Monte-Carlo integration. Alternatively, the cumbersome multi-dimensional integral can be tackled with the evidence approximation determining the most probable α^*, $\boldsymbol{b}^*$ and $\boldsymbol{h}^*$ [7].

The kernel-width distribution encodes the complexity of the model. A measure for the effective DOF (eDOF) of the model parameters, which serves as a diagnostic tool, is given by the sum of the eigenvalues of B which may be illustrated in the two limits: If the kernel widths are all very small the reconstruction is determined pointwise with eDOF=N: $f_i = h_i$. We have no correlations and obtain the conventional result with the uncorrelated entropic prior. If the kernel widths go to infinity the image is described solely by the mean of the density $\boldsymbol{h}$ with eDOF=1: $f_i = \sum_i^N h_i/N$. The large eigenvalues of B define eigenvectors which are essential for describing the data and the small eigenvalues define eigenvectors describing insignificant contributions due to noise.

The quantity that governs Occam's razor is the volume of the prior covered by the high-likelihood region picked by the data. The essential contribution is given by the integral over the hidden image.

$$\int d^N\boldsymbol{h}\, P(\boldsymbol{d}|\boldsymbol{h}, \boldsymbol{b}, I)\; P(\boldsymbol{h}|\alpha, I) \quad\propto\quad det^{-\frac{1}{2}}(B(\boldsymbol{b})^T\, H_L\, B(\boldsymbol{b}) + \frac{\alpha}{\boldsymbol{h}^*}) \tag{17}$$

The first term in the determinant is the Hessian of the logarithm of the likelihood function H_L with the kernel B and the second term arises from the Hessian of the entropy. There is a tradeoff between the likelihood function which favors small kernel widths in order to fit the data and Occam's razor which penalizes small kernel widths in order to reduce complexity.

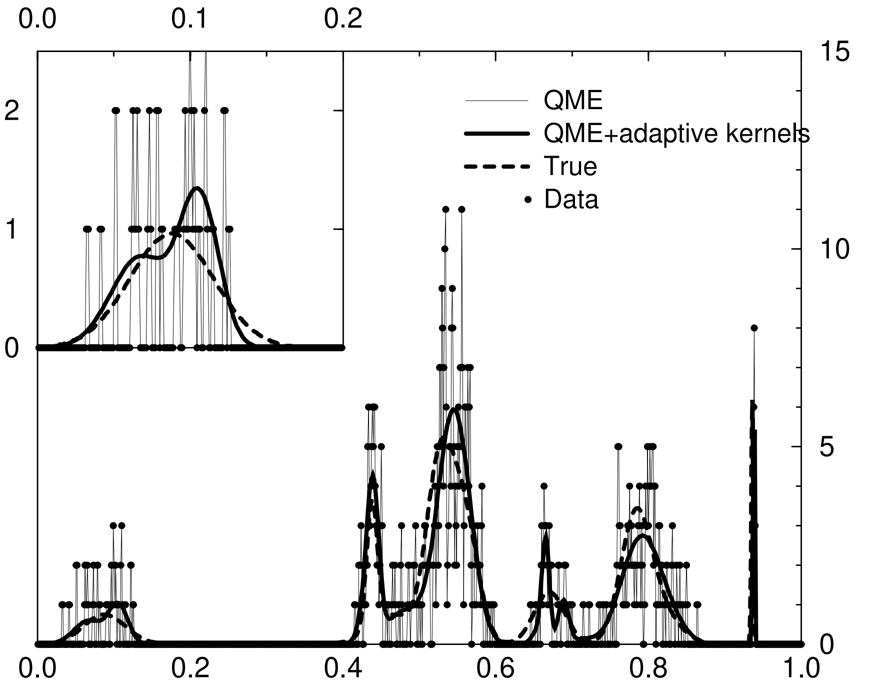

Figure 2. Density Estimation Example.

To illustrate the adaptive-kernel approach, Fig. 2 shows a mock density estimation example introduced by Silver [6], in which 800 samples are drawn *iid* from a true f (True). The data are represented as a histogram with narrow bins. Quantified Maximum Entropy (QME) fits only the sampled points. To overcome this problem Silver proposed the Quantum Entropy concept [6]. We compare his result with that of our approach. The adaptive-kernel result shows no fine structure in the left-most peak. There is only a small shoulder which reflects the double structure in the data. The peak on the right is fully resolved which is hardly obtainable with one smoothing parameter for the whole density range. The most important ingredient in avoiding ringing or noise fitting is the adaptive concept.

5. Results

Next we turn to the form-free reconstruction of the EED in a plasma. The results for a real data set consisting of 8 emission line intensities is shown in Fig. 3. The dashed line depicts a Maxwell distribution of the temperature measured with the

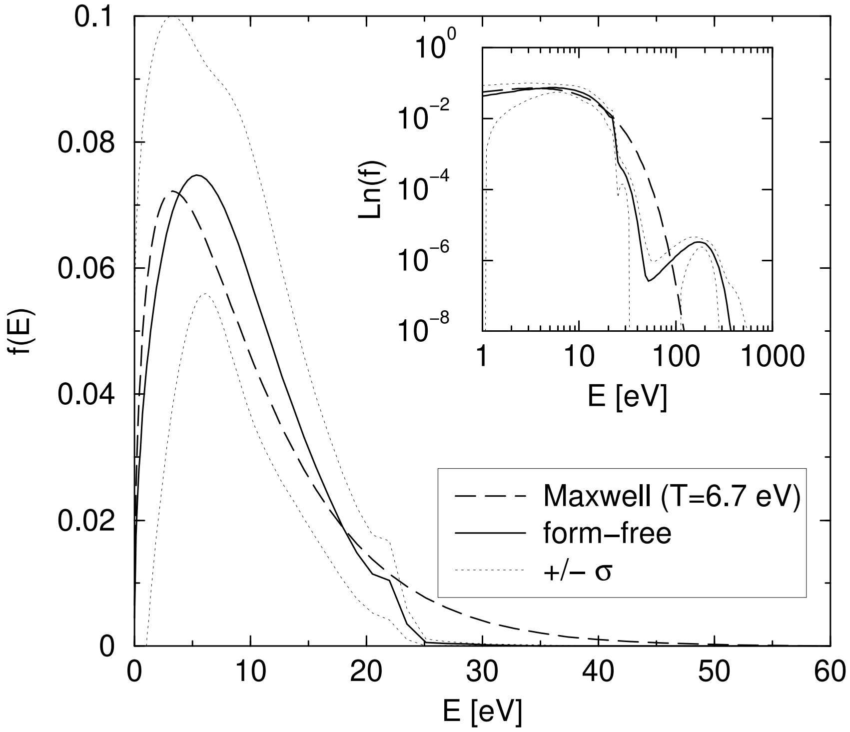

Figure 3. The dashed line depicts a Maxwell distribution of the temperature measured with the Langmuir probe. The solid line shows the form-free estimation of the EED with the adaptive-kernel approach. The steep decrease at an energy of about 20 eV arises due to a loss channel by excitation of helium. The peak at about 100-200 eV is interpreted as an accumulation of high energy electrons due to the heating mechanism.

Langmuir probe. The solid line shows the form-free estimation of the EED with the adaptive-kernel approach. The dotted line indicates the one standard deviation variance of the form-free estimation. The significant difference in the form of the distributions is quite obvious. There are two major results: The first is the steep decrease of the distribution at an energy of about 23 eV, which is even more obvious on a log-log scale. This is not a result of insufficient information in the data which would yield a smoother reconstruction compared to the Maxwell distribution. The steep decrease is driven by significant information in the data.

The threshold energies of excitation of an electron from helium in the ground state into the excited states is between 20 eV and the ionization energy of 24.6 eV. Since most helium atoms are in the ground state, the dominance of the loss channel by inelastic collisions is not surprising. At very low degrees of ionization, here less than 10^{-3}, the electron energy gained in the microwave field is essentially transferred to helium atoms by inelastic collisions. In retrospect, instead of using a Maxwellian EED a Druyvesteyn distribution fitted to the current-voltage characteristic of the Langmuir probe would do a better job.

The second new result is a small but significant peak between 100 and 200 eV which is shown in the inset. We interpret this significant additional structure as an accumulation of high energy electrons due to the heating mechanism by electron cyclotron resonance. The energetic position and amplitude of this structure is not yet understood.

Please note, that the decrease of the EED at the energy $E = 0$ is not driven by the data, but enforced by boundary conditions imposed due to the vanishing phase space in three-dimensional electron velocity space at $v = 0$, which is exact prior knowledge.

6. Summary

In summary, we have shown that ringing and noise fitting, intrinsic to ill-posed inversion problems, can be reduced drastically by the introduction of adaptive resolution into the Bayesian analysis. Occam's razor favors the smallest effective degree-of-freedom which is necessary to describe the data. The form-free estimation of the EED shows large deviations from a Maxwell distribution. There is a strong decrease where excitation of helium provides a loss channel. An additional structure arises which is interpreted as an accumulation of high energy electron as a consequence of the electron cyclotron resonance heating.

References

1. K. Behringer and U. Fantz, "Spectroscopic diagnostics of glow discharge plasmas with non-Maxwellian electron energy distributions," *Appl. Phys.*, **27**, p. 2128, 1994.
2. P. Reinke, S. Schelz, W. Jacob, and W. Möller, "Influence of a direct current bias on the energy of ions from an electron cyclotron resonance plasma," *J. Vac. Sci. Technol. A*, **10**, p. 434, 1992.
3. D. Bates, A. Kingston, and R. McWhirter *Proc. Roy. Soc. A*, **270**, p. 155, 1962.
4. J. Skilling, "Fundamentals of maxent in data analysis," in *Maximum Entropy in Action*, B. Buck and V. Macaulay, eds., p. 19, Clarendon Press, Oxford, 1991.
5. R. Fischer, W. von der Linden, and V. Dose, "On the importance of α marginalization in maximum entropy," in *Maximum Entropy and Bayesian Methods*, R. Silver and K. Hanson, eds., p. 229, Kluwer Academic Publishers, Dordrecht, 1996.
6. R. Silver, "Quantum entropy regularization," in this proceedings.
7. R. Fischer, W. von der Linden, and V. Dose, "Adaptive kernels and occam's razor in inversion problems," in *MAXENT96 - Proceedings of the Maximum Entropy Conference 1996*, M. Sears, V. Nedeljkovic, N. E. Pendock, and S. Sibisi, eds., p. 21, NMB Printers, Port Elizabeth, South Africa, 1996.