Spherical electron momentum density distribution and Bayesian analysis of the renormalization parameter in Li metal

2000 J. Phys.: Condens. Matter 12 87

(http://iopscience.iop.org/0953-8984/12/1/308)

J. Phys.: Condens. Matter **12** (2000) 87–98. Printed in the UK PII: S0953-8984(00)05556-9

Spherical electron momentum density distribution and Bayesian analysis of the renormalization parameter in Li metal

Ludwik Dobrzyński

Laboratoire de Minèralogie et Cristallographie, Universitè de Paris VI Pierre et Marie Curie, 4, place Jussieu, 75252 Paris, France
The Institute of Physics, University of Białystok, Lipowa 41, 15-424 Białystok, Poland
and
The Soltan Institute for Nuclear Studies, 05-400 Otwock-Świerk, Poland

Received 28 June 1999

Abstract. The Bayesian analysis of the spherical part of the electron momentum density was carried out with the goal of finding the best estimation of the spherically averaged renormalization parameter, z, quantifying the discontinuity in the electron momentum density distribution in Li metal. Three models parametrizing the electron momentum density were considered and nuisance parameters integrated out. The analysis show that the most likely value of z following from the data of Sakurai *et al* is in the range of 0.45–0.50, while 0.55 is obtained for the data of Schülke *et al*. In the maximum entropy reconstruction of the spherical part of the electron momentum density three different algorithms were used. It is shown that all of them produce essentially the same results. The paper shows that the accurate Compton scattering experiments are capable of bringing information on this very important Fermiological aspect of the electron gas in a metal.

1. Introduction

One of the important Fermiological characteristics of metallic systems is the influence of electron–electron correlation on the occupation number function. This function is neither easy to calculate nor to obtain experimentally. However, in addition to the topology of the Fermi surface, it is the crucial characteristics of the electron gas in a real metal, so an experimental information on it is very needed. The measurements of the so-called Compton profiles, defined as

$$J(p_z) = \int\int n(p_x, p_y, p_z)\,\mathrm{d}p_x\,\mathrm{d}p_y \tag{1}$$

where $n(\boldsymbol{p})$ is the three-dimensional electron momentum density distribution, bring information on the topology of the Fermi surface and on the behaviour of this function at the Fermi momenta. This information, however, is not very easy to obtain directly from the experiment for two reasons. Firstly, the number of crystallographic directions along which the measurements are usually carried out is not too large. This makes good reconstruction of the three-dimensional $n(\boldsymbol{p})$ function very difficult. Secondly, the energy (or momentum) broadenings which are due to the finite resolution function of the instrument smear out the discontinuities. If in addition, the expected discontinuities are direction dependent, situation becomes truly complicated. We should also stress that $n(\boldsymbol{p})$ function is not identical with the occupation number function but one can obtain the latter by using the so-called Lock–Crisp–West method (Lock *et al* 1973).

0953-8984/00/010087+12$30.00

In this paper we discuss the case of Li metal for which high-quality experimental data exist (Sakurai *et al* 1995, Schülke *et al* 1996). Lithium is very important as the simplest of metals. Simultaneously it serves as a good test case because the anisotropy of the electron momentum density is very small. Therefore the Compton scattering data can be spherically averaged in order to obtain the spherical Compton profile, and the reconstruction of $n(\boldsymbol{p})$ may be essentially reduced to the one-dimensional problem:

$$J(q) = 2\pi \int_q^\infty n(\boldsymbol{p})\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}. \tag{2}$$

If $J(q)$ represents a true Compton profile, the experimentally measured one is described by the convolution of $J(q)$ with the resolution function $R(q - q_0)$ of the instrument:

$$J_{\mathrm{exp}}(q_0) = \int R(q - q_0)J(q)\,\mathrm{d}q. \tag{3}$$

The resolution function is usually assumed to be Gaussian. In the case of the data collected by Sakurai *et al* (1995), treated in this paper, the FWHM of this function was 0.12 a.u. The mathematical, inverse problem consists in obtaining possibly precise information about the $n(\boldsymbol{p})$ function from equations (3) and (2). In order to achieve this goal we apply the maximum entropy method (Dobrzyński and Holas 1996).

In the course of the preliminary one-dimensional reconstructions by Dobrzyński *et al* (1997) it was noticed that the experimental data indicate that the renormalization parameter, z, is not smaller than 0.2. More recent three-dimensional reconstructions were pointing to the value over 0.4 (Dobrzyński, to be published). These values should be confronted with the results of the original analysis in the paper by Schülke *et al* (1996), which brought $z = 0.1 \pm 0.1$, theoretical value close to unity (Tanaka *et al* to be published) and the directionally averaged value close to 0.3 which could be inferred from the paper by Kubo (1997).

In order to obtain better insight into the situation we decided to carry out first of all the so-called Bayesian analysis of the probability of finding a given value of z-parameter in the experimental data. The paper is organized as follows. In the next section the very basis of the Bayesian analysis is recapitulated. Next, in section 3, we discuss the algorithms used. The models and results of the analysis are presented in section 4 followed by the conclusions.

2. Bayesian analysis

There exists a number of papers dealing with the very basis of the Bayesian analysis. We shall quote and recapitulate here one of the most recent papers by Gerhardt *et al* (1998), which contains most of the information needed for the purpose of our paper, and a little older one by Sivia *et al* (1993), which serves as an introduction to the Bayesian model selection problem.

Let us have the data D_k, $k = 1, \ldots, N_{\mathrm{exp}}$, which are to be described by a positive definite and additive distribution function $\{f_n\}$, $n = 1, \ldots, N_p$ through a linear relationship:

$$D_k = \sum_{l=1}^{N_p} r_{kl} f_l + \sigma_k \tag{4}$$

where the r_{kl} matrix contains already information on the model and resolution function of the instrument, and σ_k denotes the standard deviation of the kth experimental point.

For the distribution $\{f_k\}$, the conditional probability of obtaining the set $\{D_k\}$ of experimental data is

$$P(\boldsymbol{D} \mid \boldsymbol{f}) = \prod_k (2\pi D_k)^{-1/2} \exp(-\chi^2/2) \tag{5}$$

where

$$\chi^2 = \sum_k (D_k - G_k)^2/\sigma_k^2 \tag{6}$$

is the usual misfit function with G_k describing the calculated value:

$$G_k = \sum_{l=1}^{N_p} r_{kl} f_l. \tag{7}$$

Clearly, the least-squares-fit routines derive from maximizing probability (5). Now, one can also ask about the conditional probability of having the distribution $\{f\}$ given certain model $\{m\}$ and the regularization parameter α. This is equal to

$$p(\boldsymbol{f} \mid \boldsymbol{m}, \alpha)| = \prod_i (2\pi f_i/\alpha)^{-1/2} \exp(\alpha S) \tag{8}$$

where the information entropy, S, is given by

$$S = -\sum_k f_k \ln(f_k/m_k). \tag{9}$$

It is tacitly assumed that both the distributions, $\{m\}$ and $\{f\}$, are normalized to the same value.

Using the Bayes theorem, the conditional (posterior) probability of obtaining $\{d\}$ and $\{f\}$ given $\{m\}$ and α, is

$$p(\boldsymbol{D}, \boldsymbol{f} \mid \boldsymbol{m}, \alpha) = p(\boldsymbol{D} \mid \boldsymbol{f}, \boldsymbol{m}, \alpha) p(\boldsymbol{f} \mid \boldsymbol{m}, \alpha). \tag{10}$$

If in addition it is demanded that the distributions of interest are normalized, say, to Z, than it follows from equations (8) and (9) that the probability $p(D, f \mid m, \alpha)$ is proportional to $\exp(L)$, where

$$L = -\chi^2/2 + \alpha S + \beta\left(\sum f_i - Z\right) \tag{11}$$

and α and β can be treated as undetermined Lagrange multipliers. Maximizing Lagrangian L with respect to the searched distribution $\{f\}$, and demanding that the sum over f is Z, results in the equation

$$f_i = Z m_i \exp(\Delta_i/\alpha) \Big/ \sum_j m_j \exp(\Delta_j/\alpha) \tag{12}$$

where

$$\Delta_i = -\partial\chi^2/\partial f_i = \sum_{k=1}^{N_{\exp}} (D_k - G_k) r_{ki}/\sigma_k^2. \tag{13}$$

Solution of strongly nonlinear equations (12) lies in the heart of computational problems. The proper choice of the multiplier α is also not trivial. As discussed in (Gerhardt *et al* 1998), the optimal choice is

$$\alpha = \sum m_i \Delta_i^2 \Big/ \sum (f_i - m_i)^2/m_i. \tag{14}$$

Assume now that among various distributions $\{f\}$, a distribution $\{\hat{f}\}$ maximizes Lagrangian (11). Assuming next that around this solution the Lagrangian has a quadratic form:

$$L(\boldsymbol{f}) \approx L(\hat{\boldsymbol{f}}) + 0.5 \sum_{i,j} (\partial^2 L/\partial f_i \partial f_j)(f_i - \hat{f}_i)(f_j - \hat{f}_j) \tag{15}$$

one can integrate out the distribution $\{f\}$ from $P(\boldsymbol{D}, \boldsymbol{f} \mid \boldsymbol{m}, \alpha)$ and obtain the value of conditional probability $p(\boldsymbol{D} \mid \boldsymbol{m}, \alpha)$, which will quantify the degree of belief we should have in the model $\{m\}$. Such calculations allow for relatively easy model selection.

3. Algorithms

The commercially distributed algorithm from the group of Cambridge was essentially described in Skilling and Bryan (1984; see also Bryan 1990). The algorithm is robust and mathematically stable. However, its cost is prohibitive for many groups wishing to use the maximum entropy method. Therefore less advanced, yet efficient algorithms, are constructed and described in the literature.

Probably the simplest way of solving equations (12) was suggested by Sakata *et al* (1993), who elaborated so-called MEED package. The method consists in calculating Δ values, equation (13), for given $\{m\}$, then calculating $\{f\}$ from equation (12), and treating this set of $\{f\}$ as a new model $\{m\}$. At every such iteration step one is controlling the value of χ^2. When it falls below the value of N_{exp} (or 1 if the normalized χ^2 values are used), the program is terminated. The Sakata *et al* method (MEED package) is easy to program and quite efficient. The program convergence is also attained relatively easily. The main objections come from the argument that during the calculations one is changing the initial model, so the entropy is not really maximized with respect to this model. The experience with using Cambridge and MEED (or MEED-type) algorithms in selected cases, see e.g. the paper by Dobrzyński *et al* (1996), tells us that the results are not very different, so this argument may be in reality much weaker than it seems at the beginning. We shall call this algorithm A1.

Recently, Gerhardt *et al* (1998) suggested the linearization of equation (12) and searching the solution for the set of $\{\Delta\}$ instead of $\{f\}$. Generalizing a little the formulae given in their paper on the case when the number of experimental points N_{exp} is not equal to generally larger number of points N_p at which reconstruction is carried out, one is solving the equations:

$$\sum_{j=1}^{N_p} M_{ij}\Delta_j^{(\nu)} = C_j \tag{16}$$

where $(i, j) = 1, \ldots, N_p$ and

$$M_{ij} = H_{ij} f_{ij}^{(\nu-1)}/\alpha + \delta_{ij} \tag{17}$$

$$H_{ij} = \sum_{k=1}^{N_{\text{exp}}} r_{ik} r_{jk}/\sigma_k^2 \tag{18}$$

$$C_i = d_i + \sum_{k=1}^{N_p} H_{ik} f_k^{(\nu-1)}(\Delta_k^{(\nu-1)} - \alpha)/\alpha \tag{19}$$

$$d_i = \sum_{j=1}^{N_{\text{exp}}} D_j r_{ji}/\sigma_j^2 \tag{20}$$

$$f_i^{(\nu)} = m_i^{(\nu)} \exp(\Delta_i^{(\nu)}/\alpha). \tag{21}$$

Index ν numerates the iteration step. Within this notation

$$p(\boldsymbol{D} \mid \boldsymbol{m}, \alpha) = \left(\prod_k (2\pi D_k)^{-1/2} \exp[L(\hat{f})]\right) \Big/ \left(\sqrt{\det(\mathbf{I} + \alpha^{-1}[\hat{f}]\mathbf{H})}\right) \tag{22}$$

with $\mathbf{I}$ denoting the unit matrix and $[\hat{f}]$ being a diagonal matrix. The algorithm of Gerhardt *et al* (1998) will be called A2.

Very similar linearization can be carried out with respect to the f_i itself. Denoting

$$\delta \boldsymbol{f} = \boldsymbol{f}^{(\nu+1)} - \boldsymbol{f}^{(\nu)} \tag{23}$$

one obtains again equation (16) with Δs substituted by $\delta \boldsymbol{f}$ and the following expressions for M_{ij} and C_i:

$$M_{ij}^{(\nu)} = \delta_{ij} + (1/\alpha) H_{ij} m_i \exp\left\{\left(d_i - \sum_k H_{ik} f_k^{(\nu)}\right) \Big/ \alpha\right\} \tag{24}$$

$$C_i^{(\nu)} = m_i \exp\left\{\left(d_i - \sum_k H_{ik} f_k^{(\nu)}\right) \Big/ \alpha\right\} - f_i^{(\nu)}. \tag{25}$$

Similarly to A2, the value of α can be chosen according to equation (14). This algorithm will be denoted as A3.

4. The results

4.1. Models

At first we have been interested in the strength of the models which could describe the experimental data (Sakurai *et al* 1995) for lithium. In all the analysis the spherically averaged data were used. The spherical averaging was carried out in a standard way by calculating weighted average of the Compton profiles with weights equal to the multiplicity of a given crystallographic direction.

Three models of the electron momentum density distribution were tried. The first one, denoted by M1, postulates

$$n(x) = \begin{cases} a_1(1 - a_2 x^2) - (18/55) T x^n & \text{for } x \leqslant 1 \\ a_3 \,\mathrm{e}^{-a_4 x} + T x^{-n} & \text{for } x > 1 \end{cases} \tag{26}$$

where

$$x = p/p_F. \tag{27}$$

The coefficient (18/55) was chosen for historical reasons and is not very essential as further analysis shows up. If equation (26) described occupation number function for a uniform electron gas the preferred value of the exponent n would be 8 (see Takada and Yasuhara 1991). Although the model contains six parameters, positive by definition, they are bounded by the requirement that the integral over $n(x)$ should be equal to 1. This requirement also imposes the exponent n to be larger than 3. In reality, we are interesting in the parameter z, i.e. the value of the discontinuity at $x = 1$. One can easily see from (26) that

$$z = a_1(1 - a_2) - (73/55) T - a_3 \,\mathrm{e}^{-a_4}. \tag{28}$$

Obviously, the value of z should be larger than 0 and smaller than 1. The physical space of parameters is also limited by the requirement that $n(1)$ should not be smaller than z.

The model M2 is taken directly from the paper (Schülke *et al* 1996):

$$n(x) = \begin{cases} b - 0.5(b - z) x^n & \text{for } x \leqslant 1 \\ 0.5(b - z) x^{-n} & \text{for } x > 1. \end{cases} \tag{29}$$

Again, the value $n = 8$ was originally used. This followed from the paper by Takada and Yasuhara (1991) who calculated the influence of electron–electron correlations on the occupation number function in uniform electron gas within so called jellium model. In light of the paper by Kubo (1997) we know that this may be a crude approximation only. In addition, in reconstructing the electron momentum density function we might have some factors difficult to control which can effectively change the rate of the disappearance of $n(x)$ with x above the Fermi momentum. Therefore the use of n as a free parameter (however larger than 3) seems

justified. As previously, we should take care that the integral $n(x)$ over whole available space should be equal to one.

Finally, the model M3 is intermediate between the two models described above and reads

$$n(x) = \begin{cases} b(1 - cx^2) - 0.5(b - z)x^n & \text{for } x \leqslant 1 \\ 0.5(b - z)x^{-n} & \text{for } x > 1. \end{cases} \tag{30}$$

4.2. Simulations

First we should like to comment on the feasibility of z-values obtained from the studies of the likelihood function integrated over nuisance parameters. In order to have better insight into this problem we have chosen model M2 as the base and simulated the experiment in which the momentum density distribution was described by the function (29). Two parameters, z and n, characterize the simulated spectrum. This spectrum was convoluted with the resolution function of 0.12 a.u. (as in the real experiment). Finally, Gaussian noise at the level of 0.5% in the peak was imposed. Next, using the method described in this paper, the $P(z)$ distributions were calculated. The results are presented in figures 1 and 2. Our experience was that in some cases one obtains a peak splitting, but even in such a case the splitting takes place around the proper value of z. It can be inferred from figure 2 that the method allows for the estimation of z with the accuracy of roughly 10%.

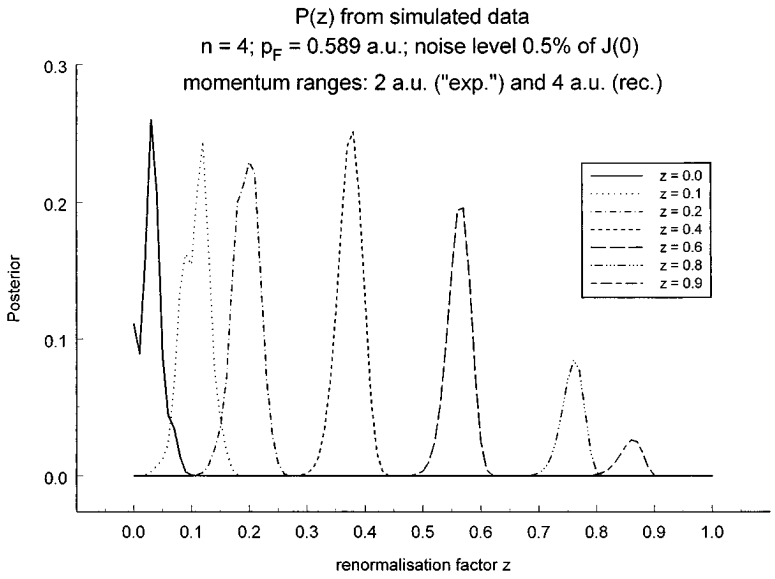


Figure 1. The distributions $P(z)$ obtained for various simulated spectra. The parameters used are shown in the figure.

Obviously, one can always argue that a real metal like lithium may have a z-parameter which is direction dependent due to various solid-state effects. It is very difficult to obtain a three-dimensional picture of this dependence unless extremely accurate data are collected for many crystallographic directions. Nevertheless we see that such a sensitive parameter like z can be estimated—even through its 'average' value—with reasonable accuracy with the help of a relatively simple procedure. In fact we are unaware of other methods which can visualize such a parameter in a more direct way.

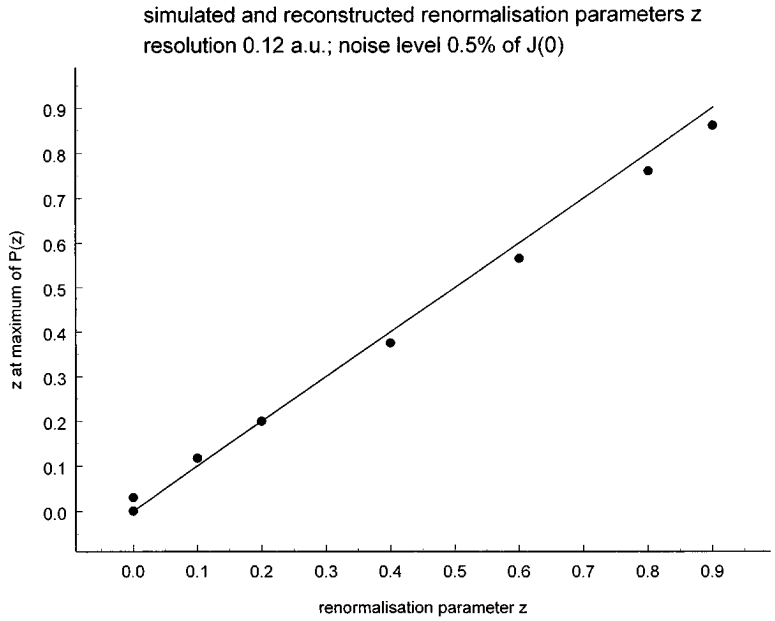


Figure 2. The value of z-parameter at the maximum of $P(z)$ function versus the values used in simulated data.

4.3. Analysis of experimental results

Figure 3(a) shows the results of calculations of the likelihood function $\exp(-\chi^2/2)$ for the model M1, integrated over all but z- and n-parameters. In all the results presented below we have been taking into account the instrumental resolution function with FWHM $= 0.12$ a.u. and $p_F = 0.589$ a.u. We see that almost irrespectively of n, the value of z at which the maximum is attained is close to 0.1–0.15. Integrating out the parameter n results in the distribution shown in figure 3(b), which exhibits a clear maximum at $z = 0.15$. Two other models, however, bring different values of z, see figures 4(a) and (b). Here the value of z is close to 0.45–0.5, so the question arises of which range of z-parameter is the most likely. In order to answer this question we have to normalize the probabilities properly and when this is done we obtain the following result:

$$P_1 : P_2 : P_3 = 0.000\,082 : 0.0104 : 0.0028. \tag{31}$$

This clearly indicates that $z = 0.45$ is favoured. In fact the analysis shows that there is a strong penalty for using many parameters in the model. For example, in the model M3 the minimum χ^2 value is as low as 3.4, while in the model M2 it reads 4.4. Nevertheless we see that the probability of M2 is almost four times larger than of model M3 which means that the penalty factor for using only one parameter more was more than 6.

One can also ask whether the chosen value of Fermi momentum p_F does not play a role. Indeed, changing p_F, the values of z at which the maxima of $P(z)$ appear change. We can, however, treat p_F as an another nuisance parameter and integrate it out. As a result, staying within the model M2, we obtain the result displayed in figure 5, which shows $z = 0.47$ as the most likely value.

This value disagrees with $z = 0.1 \pm 0.1$ obtained in Schülke *et al* (1996) and is also larger than the value about 0.3 which can be deduced from the paper by Kubo (1997). In order to

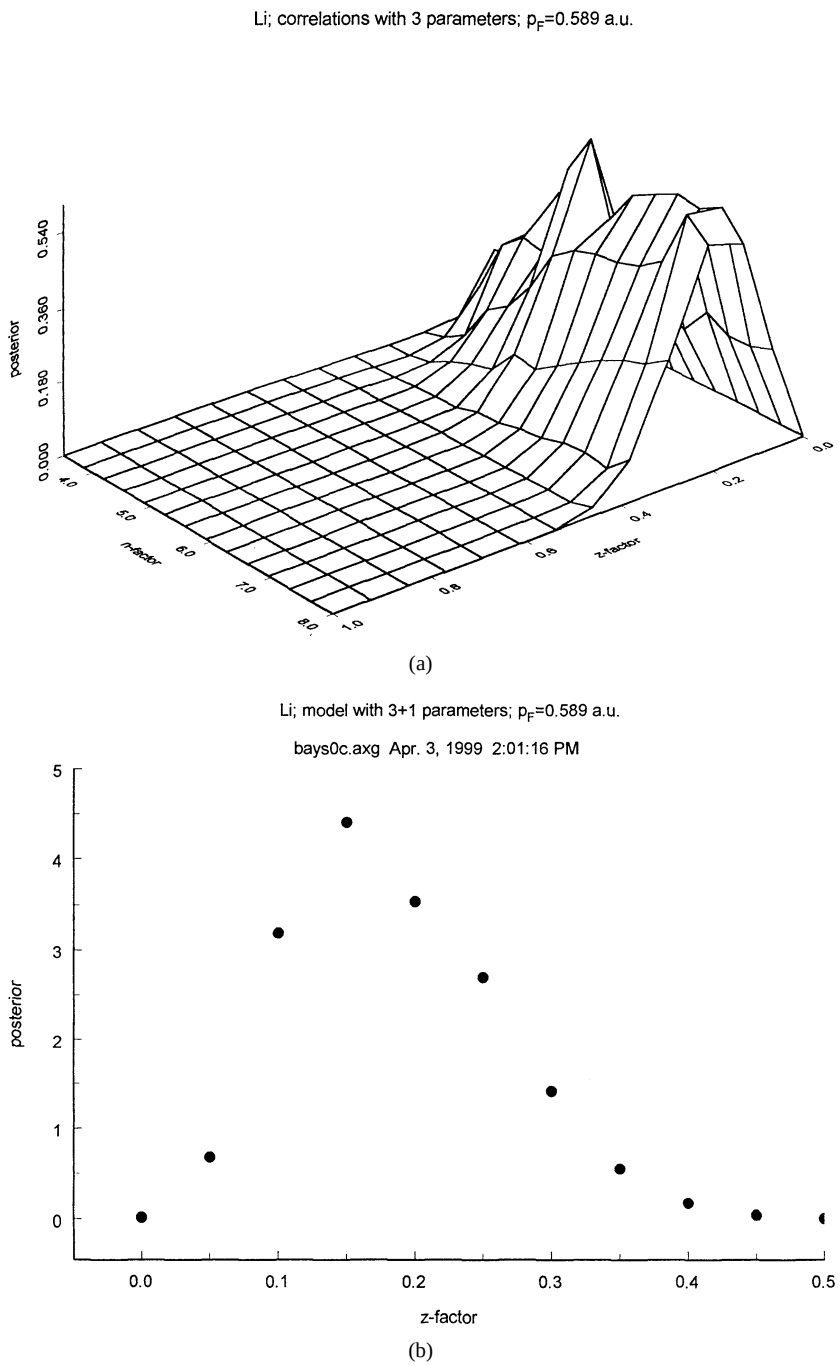
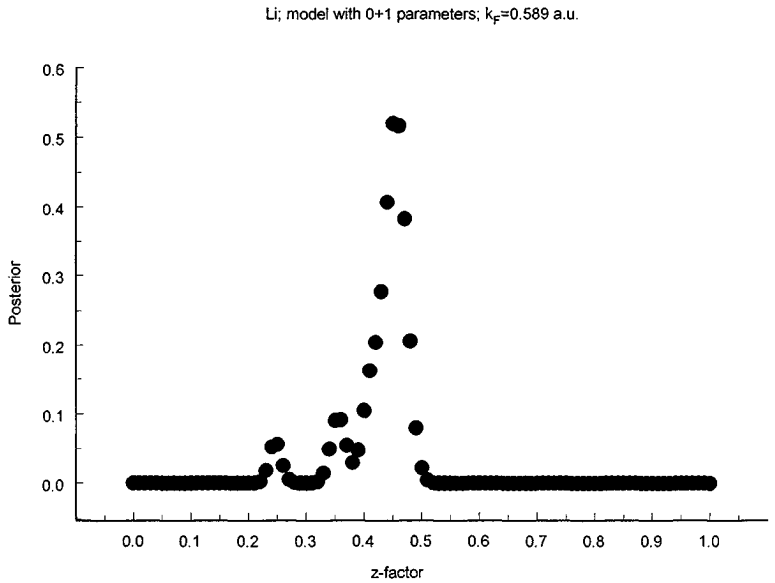
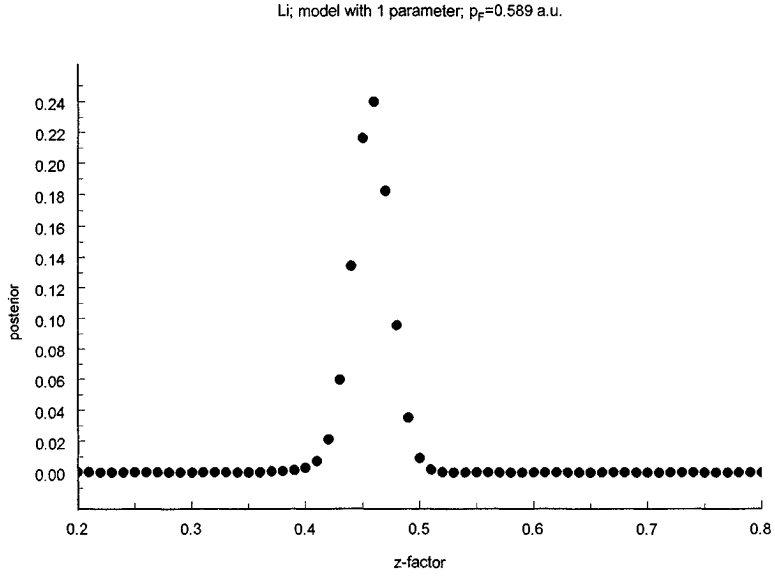


Figure 3. (a) Likelihood function for the model M1 after marginalization with respect to all but z and n parameters. (b) The same as figure 3(a) but after marginalization also with respect to n.

check whether the lithium data obtained by the German group, which are slightly different than the ones given by the Japanese–American team, are not leading to essentially different values of z, we carried out analysis of the German data. We adopted the resolution width of



(a)



(b)

Figure 4. (a) Same as figure 3(b) for the model M2. (b) The same as figure 3(b) for the model M3.

0.17 a.u. (Döring, private information). Although we expected to find a value of z close to 0.1, as published in Schülke *et al* (1996), we found a narrow distribution of possible z with the maximum at 0.55 and the width at half maximum of 0.05, see figure 6. So the renormalization parameters given by the two experiments are almost the same and it is possible that the use of the $n = 8$ value could be critical for the estimation of z. In order to give more precise answer yet staying within the Bayesian approach, we would need to know a little more, e.g. the strict range of possible values of some of the parameters appearing in the models. At this

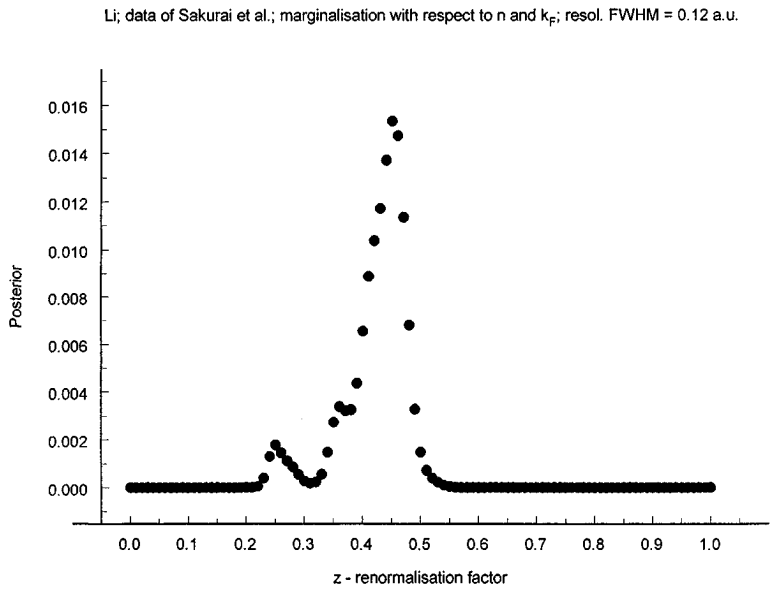


Figure 5. Same as figure 4(a) with additional marginalization with respect to p_F.

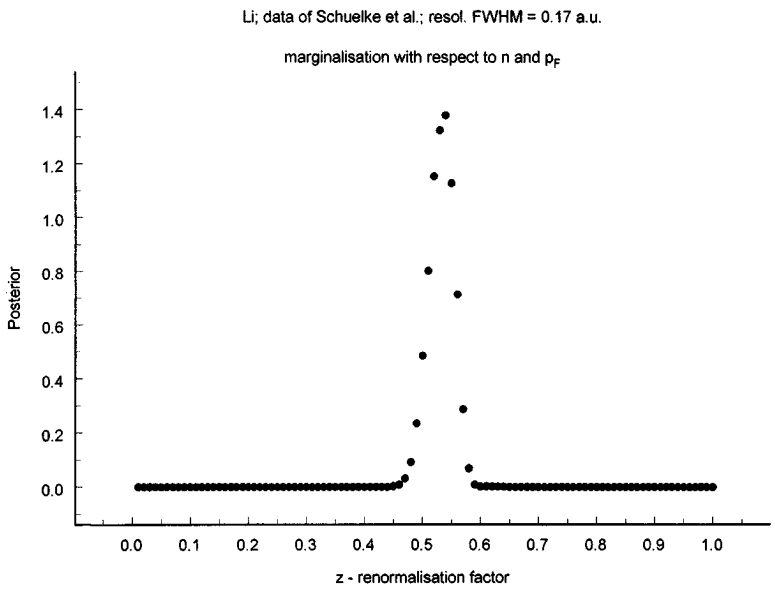


Figure 6. Same as figure 5 for the data of Schülke *et al* (1996).

stage of analysis, we should say that $z = 0.45$–0.55 is the most likely range of variation of this parameter.

In the final step of our analysis we checked the results of the reconstruction of the spherical part of the electron momentum density distribution by the maximum entropy method using the algorithms described in section 3. It has been shown by Dobrzyński and Bansil (unpublished) that using the prior given by the model M3 with $z = 0.4$ results in the density distribution which exhibits about 10% larger values at low momenta with correspondingly lower values at

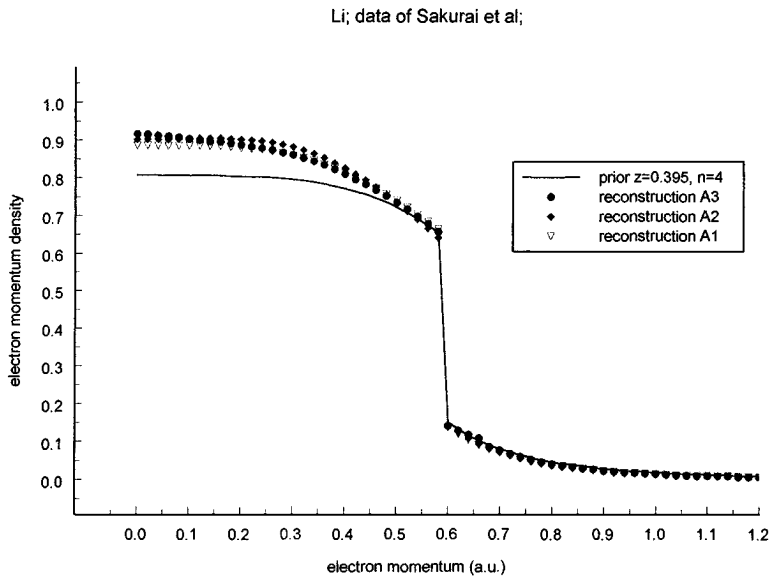


Figure 7. Reconstructed spherical electron momentum density distribution for lithium using different algorithms.

larger momenta. The ubrupt jump at the Fermi momentum stays, however, intact. Because the algorithm (A1) used there was just a variation of the algorithm of Sakata *et al*, this time we used first the algorithm A3, which proved to be much faster than the algorithm A1. In the case of the analysed data about 60 iterations were needed to achieve full convergence of the program. Similar calculations carried out with the use of algorithm A2 were also fast. Results obtained with help of all three algorithms are displayed in figure 7. We see that although minute differences exist, the most essential result for the value of the renormalization parameter is well preserved.

5. Conclusions

Bayesian analysis has been carried out in order to establish the most likely value of discontinuous jump in the electron momentum density distribution at the Fermi momentum. This value should be close to the value of the renormalization parameter, z. As a test case the results of the Compton profiles of lithium measured in Sakurai *et al* (1995) were taken to the analysis. Three different models parametrizing function $n(x)$ where $x = p/p_F$, were used. As a result it was found that the most likely value is in the range of 0.45 to 0.50, while the data of Schülke *et al* (1996) point to $z = 0.55$. In order to change these results one would need to use some extra knowledge about the parameters entering the model or to propose better models. Therefore the success of this paper merely consists in showing that the renormalization parameter can be obtained from the Compton experiments carried out with the resolution of 0.12 a.u. and even 0.17 a.u.

Independently it was checked that the simple algorithm A1, used in the MEED package, is producing results very close to the ones produced by other two algorithms tried in this paper. This gives more confidence not only in the A1 type algorithms but also in the result concerning the value of the z-parameter, which turned out to be not affected by the process of reconstruction of the spherical part of the electron momentum density in lithium.

Acknowledgments

The author wishes to thank University of Paris VI who ensured excellent conditions of work. Professor Genevieve Loupias and Dr Christophe Bellin are particularly thanked for their hospitality and many useful discussions during the author's stay in Paris. Dr Robert Papoular from Laboratoire Léon Brillouin at Saclay deserves special thanks as without his help and numerous advices this paper would never appear. I am also grateful to Professor Nobuhito Shiotani and Dr Yoshikazu Tanaka as well as to Professor Winifried Schulke and Dr Gordon Doering for providing me with their original data.

References

Bryan R K 1990 *Eur. Biophys. J.* **18** 165
Dobrzyński L and Holas A 1996 *J. Nucl. Instrum. Methods* A **383** 589
Dobrzyński L, Papoular R and Sakata M 1996 *J. Phys. Soc. Japan* **1** 255
Dobrzyński L, Tanaka Y and Sakurai Y 1997 *Mater. Sci. Forum* **255–257** 150
Gerhardt U, Marquardt S, Schroeder N and Weiss S 1998 *Phys. Rev.* B **58** 6877
Kubo Y 1997 *J. Phys. Soc. Japan* **66** 2236
Lock D G, Crisp V H C and West R N 1973 *J. Phys. F: Met. Phys.* **3** 561
Sakata M, Uno T, Takata M and Howard Ch J 1993 *J. Appl. Crystallogr.* **26** 159
Sakurai Y, Tanaka Y, Bansil A, Kaprzyk S, Stewart A T, Nagashima Y, Hyodo T, Nanao S, Kawata H and Shiotani N 1995 *Phys. Rev. Lett.* **74** 2252
Schülke W, Stutz G, Wohlert F and Kaprolat A 1996 *Phys. Rev.* B **54** 14 381
Sivia D S, David W I F and Knight K S 1993 *Physica* D **66** 234
Skilling J and Bryan R K 1984 *Mon. Not. R. Astron. Soc.* **211** 111
Takada Y and Yasuhara H 1991 *Phys. Rev.* B **44** 7879
Tanaka Y, Bansil A, Kaprzyk S, Sakurai Y, Stewart A T and Shiotani N to be published