Bayesian deconvolution and analysis of photoelectron or any other spectra: Fermi-liquid versus marginal Fermi-liquid behavior of the 3*d* electrons in Ni

U. Gerhardt
Physikalisches Institut der Johann Wolfgang Goethe-Universität, Robert-Mayer-Straße 2-4, D-60054 Frankfurt am Main, Germany

S. Marquardt
Ruhrgas AG, D-45138 Essen, Germany

N. Schroeder
Standard Application Software AG, D-69190 Walldorf (Baden), Germany

S. Weiss
Physikalisches Institut der Johann Wolfgang Goethe-Universität, Robert-Mayer-Straße 2-4, D-60054 Frankfurt am Main, Germany
(Received 19 May 1997; revised manuscript received 13 May 1998)

We present a simple and effective iterative deconvolution of noisy experimental spectra *D* broadened by the spectrometer function. We show that this ''iterative Bayesian deconvolution'' is closely related to the more complex ''Bayesian analysis,'' also known as the quantified maximum-entropy method. A model *m* of the true spectral function is needed in both cases. The Bayesian analysis is the most powerful and precise method to relate measured spectra *D* to the corresponding theoretical models *m* via the respective probabilities, but two grave conceptual problems together with two severe technical difficulties prevented widespread application. We remove these four obstacles by (i) demonstrating analytically and also by computer simulations that the most probable deconvolution $\hat{a}$ obtained as a by-product from the Bayesian analysis gets closer to the true spectral function as the quality of *m* increases, (ii) finding it equivalent but vastly more efficient to optimize the parameters contained in a given model *m* by the usual least-squares fit between *D* and the convolution of *m prior* to the Bayesian analysis instead of using the Bayesian analysis itself for that purpose, (iii) approximating the convolution by a summation over the energies of the *n* data points only, with the normalization of the spectrometer function chosen to minimize the errors at both edges of the spectrum, and (iv) avoiding the severe convergence problems frequently encountered in the Bayesian analysis by a simple reformulation of the corresponding system of *n* nonlinear equations. We also apply our version of the Bayesian analysis to angle-resolved photoelectron spectra taken at normal emission from Ni(111) close to the Fermi energy at about 12 K, using two different physical models: Compared with the marginal Fermi liquid, the Fermi-liquid line shape turns out to be about 10^4 times more probable to conform with the observed structure of the majority and minority spin peaks in the low-photon and small-binding-energy region. [S0163-1829(98)04435-X]

I. INTRODUCTION

The deconvolution of a noisy experimental spectrum, broadened by the spectrometer function and taken in a limited range only, is an ill-posed problem, i.e., there is a broad range of spectral functions, the convolution of which all fit the measured spectrum equally well. Most of these spectral functions, however, show strong unphysical oscillations and can be weeded out by introducing smooth model spectral functions that take the evidence of the measured spectrum approximately into account. Far from being arbitrary, this analysis goes beyond the simple deconvolution since it allows us to assess the validity of different theoretical models in terms of the measured spectrum quantitatively.

In the first part of this paper we discuss the merits of two different deconvolution procedures by comparing their results with the corresponding values of the *true* spectral function $\tilde{a}_j=\tilde{a}(E_j)$, i.e., we start out from the data of an ''experimental spectrum'' $D_k=D(E_k)$ obtained by adding Poisson distributed noise P_k with the average magnitude $(D_k)^{1/2}$ to the convolution $G(\tilde{a})$ of $\tilde{a}$, i.e.,

$$D_k=G_k(\tilde{a})+P_k, \quad G_k=\sum_{j=1}^{n} a_j G_{jk}, \tag{1}$$

where G_k is the convolution of any a_j and G_{jk} is the normalized Gaussian spectrometer function characterized by the width parameter σ,

$$G_{jk}=R_{jk}/N_k, \quad R_{jk}=\exp[-(E_j-E_k)^2/2\sigma^2],$$

$$N_k=\sum_{h=1}^{n} R_{hk}. \tag{2}$$

The first method is our own iterative deconvolution, which dramatically improves upon the somewhat similar method proposed by van Cittert[1] since it is closely related to the second method, which is the complete Bayesian spectrum analysis given by Skilling.[2] Application of this powerful Bayesian analysis to, e.g., photoelectron or inverse photoelectron spectra has been sparse[3,4] since the role of the model has not been fully appreciated and since the method seemed to preclude the optimization of more than one or two param-

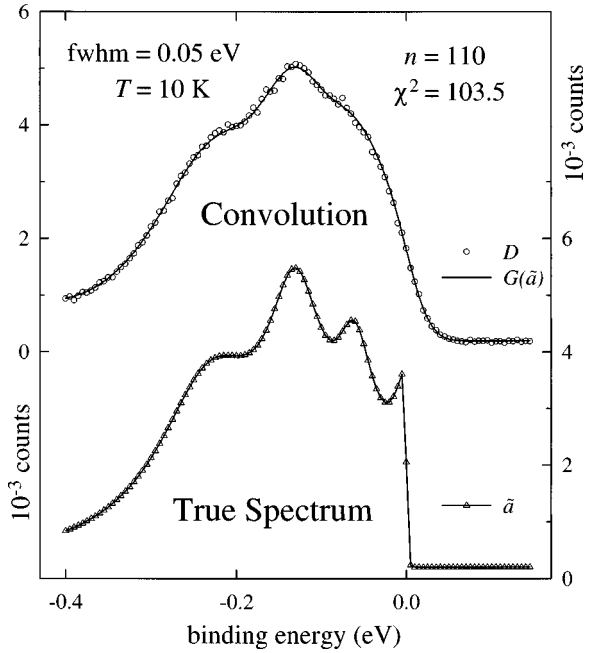

FIG. 1. The circles of the upper curve give the ''experimental spectrum'' D for which different deconvolutions will be shown in Figs. 2–8. It has been generated by convoluting the true spectral function $\tilde{a}$ (the triangles in the lower part) with the spectrometer function and adding Poisson distributed noise to this convolution (solid line in the upper part). The circles of this experimental spectrum and the triangles of this true spectral function are also shown in Figs. 2–8.

eters contained in the model, a situation that we set out to rectify. We also simplify the Bayesian analysis considerably by introducing two important technical improvements.

In the second part of the paper we apply the Bayesian analysis[2] including our own amendments to the actual angle and energy resolved photoelectron spectra of Ni(111), taken at about $T=12$ K near the Fermi energy E_F. Our aim is to distinguish between the Fermi-liquid and the marginal Fermi-liquid[5,6] behavior of the interacting $3d$ electrons, respectively.

II. ITERATIVE BAYESIAN DECONVOLUTION

The lower part of Fig. 1 shows the true spectral function $\tilde{a}$ as triangles. It consists of a superposition of four Lorentzians with different intensities and energy positions, truncated by the Fermi-Dirac function with $T=10$ K, and a small constant presenting the dark count. One of the Lorentzians has its center above the Fermi energy $E_F=0$. The width of the Lorentzians increases quadratically with the binding energy. We use Eq. (1) to calculate the convolution, which is an approximation since we replace the integral by a sum and also because we carry this sum over n data points only. We do so because we select the experimental spectrum itself as the first starting model of the spectral function, as proposed originally by van Cittert.[1] The normalization chosen in Eq. (2) minimizes the errors at the edges of the spectrum caused by this restriction since at least for $\tilde{a}=$const the correct result, which is $G(\tilde{a})=\tilde{a}$ in this case, is obtained also up to the edges. In fact, as also expected from symmetry considerations, using G_{jk} of Eq. (2) in Eq. (1) results in a convolution with only small deviations from the true convolution at the edges if the expansion of a around the edges contains even powers only. If a varies linearly at the edges these errors will be larger, but they can be reduced by starting and stopping the measurement in regions where the slope of D is small. The upper part of Fig. 1 shows $G(\tilde{a})$ as the solid line, calculated using Eqs. (1) and (2), with a 0.05 eV full width at half maximum (FWHM) of the spectrometer function R_{jk}, and D as circles. The probability distribution[7]

$$F(\chi^2)\propto\exp(-\chi^2/2)(\chi^2)^{(n/2-1)} \tag{3}$$

of the quantity

$$\chi^2=\sum_{k=1}^{n}[(D_k-G_k)^2/D_k] \tag{4}$$

has its maximum at $n-2$ for $G=G(\tilde{a})$, where $n-2=108$ in the example. This is reasonably close to the actual value of $\chi^2(\tilde{a})\approx 103.5$ for the particular noise spectrum of Fig. 1. Actually, $\chi^2(\tilde{a})$ must be close to n owing to Eq. (4) since $[D_k-G(\tilde{a})]^2\approx\sigma_k^2=D_k$.

We now try to recover $\tilde{a}$ from D using an iterative deconvolution: The values $a_i^{(\nu)}$ obtained in the step ν are used to get $a_i^{(\nu+1)}$ from

$$a_i^{(\nu+1)}=a_i^{(\nu)}\exp(\Delta_i/\alpha), \tag{5}$$

where Δ_i is some reasonable measure of the difference between D and $G^{(\nu)}=G(a^{(\nu)})$, containing the $a_j^{(\nu)}$ according to Eq. (1). This is the basic equation that also governs the full Bayesian analysis to be discussed later; it has the virtue to fulfill the obvious requirement that $a_i\geqslant 0$ for all i. The parameter α can be used to regulate the speed of convergence in the iterative deconvolution; for simplicity, we set $\alpha=1$ here. However, in the Bayesian analysis α will play a crucial role, which is the reason that we show it explicitly in Eq. (5).

Computer simulation of the iterative deconvolution

What is the correct expression for Δ_i? Should we follow the early proposal by van Cittert[1] and use $\Delta_i=[D_i-G_i(a^{(\nu)}]/D_i$? Neglecting the noise, this choice would yield $\Delta_i\to 0$ for $\nu\to\infty$ or $a\to\tilde{a}$ using Eq. (5), but in reality the noise component of D_i is transferred to a_i at each step of the iteration, giving rise to

$$a_i=\tilde{a}_i+a_i^{\mathrm{osc}} \quad \text{with} \quad G_k^{\mathrm{osc}}=\sum_{i=1}^{n}a_i^{\mathrm{osc}}G_{ik}\approx 0 \tag{6}$$

instead, i.e., we obtain the unphysical a mentioned in the Introduction, consisting of the true $\tilde{a}$ and the oscillatory component $a^{\mathrm{osc}}\propto D-G(\tilde{a})$, which is the amplified noise spectrum of D according to Eq. (1), the amplitude of which increases as the iteration progresses. We perform several iterative van Cittert deconvolutions, using the linearized version of Eq. (5) and the experimental spectrum as the starting model, $m_u=a^{(0)}=D$, as also originally proposed by van Cittert.[1] As expected, the deconvolution approaches the form given in Eq. (6) after only a few iterations.

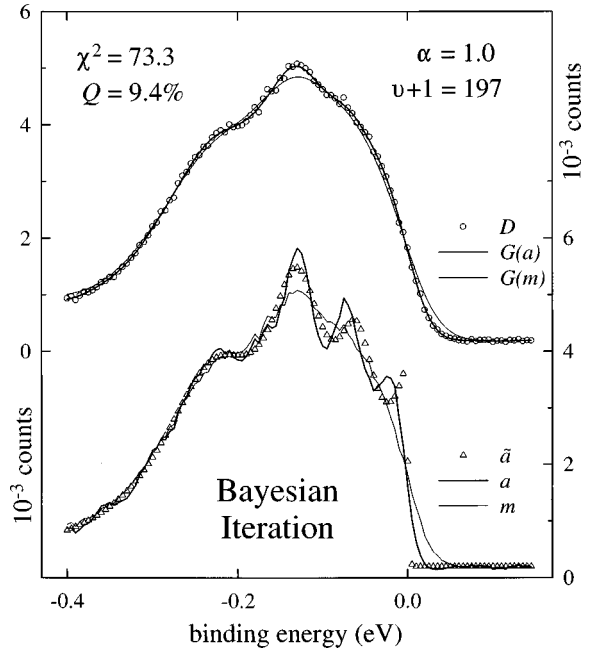

FIG. 2. Iterative Bayesian deconvolution a and $G(a)$ for $\nu+1=197$. The experimental spectrum D itself is taken as the model m. See the text for χ^2, Q, and α.

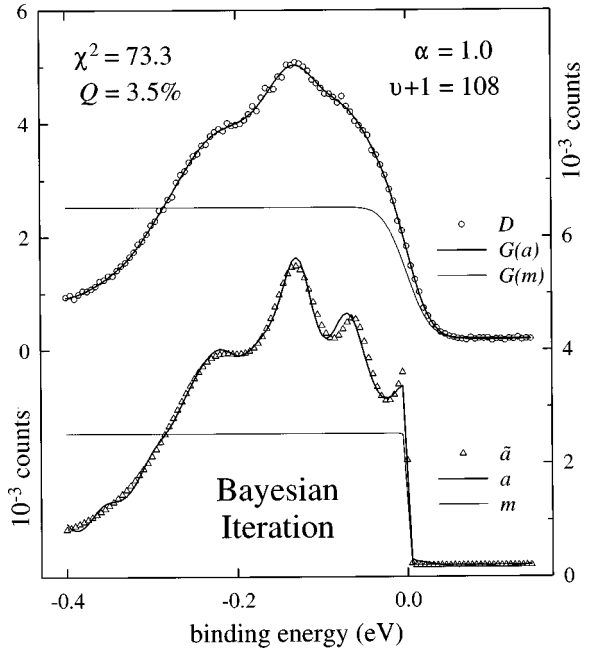

FIG. 3. Iterative Bayesian deconvolution a and $G(a)$ where the model is a constant multiplied by the Fermi-Dirac function plus a small dark count. The number of iterations $\nu+1=108$ is adjusted to obtain $\chi^2\approx 73$ as in Fig. 2.

The van Cittert results sketched above are not encouraging. However, one of us (U.G.) kept being fascinated with the basic idea. While trying to improve the method, he noticed another defect: The convolution distributes the information contained in a_i at E_i among the various D_k in the neighborhood of E_i. If one wants to improve a_i iteratively using Eq. (5), one should thus take notice of these neighboring D_k, not only of D_i as implied by the van Cittert Δ_i. The most obvious weighting factor is of course the normalized spectrometer function G_{ki} of Eq. (1). He therefore proposed to replace Δ_i in Eq. (4) by the weighted average of the relative deviations between D and $G^{(\nu)}$,

$$\Delta_i=\sum_{k=1}^{n}\frac{D_k-G_k^{(\nu)}}{D_k}\,G_{ki}\,, \tag{7}$$

which can also be interpreted as the convolution of the relative deviations between D and $G^{(\nu)}$. This substitution simultaneously reduces the severe first defect of the van Cittert method to a minor nuisance since Eq. (7) effectively averages out the noise contained in the D_k over the region centered at E_i and specified by the width of G_{ki}.

Figure 2 shows the result of the iterative deconvolution using Eq. (7) in Eq. (5) and $m_u=a^{(0)}=D$ as the starting model. The number of iterations is adjusted to obtain $\chi^2\approx 73$ to allow an easy comparison with the corresponding results of the Bayesian analysis of Sec. III. An average deviation of $Q=9.4\%$ between a and $\tilde{a}$ is obtained, where Q is defined as

$$Q=\sqrt{n\Sigma(a_i-\tilde{a}_i)^2}\Big/\sum\tilde{a}_i\,. \tag{8}$$

In comparison to a corresponding iteration using the van Cittert[1] $\Delta_i=(D_i-G_i^{(\nu)})/D_i$ for which $Q=23.6\%$ is obtained at the same $\chi^2\approx 73$, the improvement is astounding in that the structure of $\tilde{a}$ that is completely swamped by the large $a^{\rm osc}$ component in the van Cittert deconvolution is now easily recognized in a since $a^{\rm osc}$ has been decreased dramatically. The strong decrease in Q by 60% also demonstrates the increased quality of the deconvolution. To our knowledge the method leading up to Fig. 2 is the easiest and fastest way to obtain a reliable deconvolution with a low $|a^{\rm osc}/\tilde{a}|$ ratio. The procedure is also ''unbiased'' since the choice $m_u=a^{(0)}=D$ does not refer to any theoretical model of the spectral function but takes the evidence in D approximately into account.

To be sure, differences between a and $\tilde{a}$ remain: For example, in Fig. 2 the Fermi edge of a is not as steep as the true Fermi edge of $\tilde{a}$. The steepness of the edge actually increases with increasing ν, decreasing $a^{\rm osc}$ in this region, but there is a simultaneous increase of $a^{\rm osc}$ at other energies: The overall quality of the deconvolution as measured by Q of Eq. (8) thus does not change drastically over a fairly large range of ν. Improvement is possible if additional knowledge about $\tilde{a}$ is available. This is usually the case in photoelectron spectroscopy where the position of E_F is routinely determined with high precision from a separate spectrum of, e.g., a clean polycrystalline Pt sample kept at the same temperature as, and in electrical contact with, the crystal to be investigated. In the next example we therefore use as the model a constant spectral function, multiplied with the Fermi-Dirac function, with $E_F=0$ eV and $T=10$ K, plus a constant dark count. Below E_F this model has actually a strong negative bias since $G(m)$ clearly contradicts the evidence in D. Nevertheless, the iterative Bayesian deconvolution shown in Fig. 3, which is again carried to $\chi^2\approx 73$, follows $\tilde{a}$ quite closely, as documented by $Q=3.5\%$ only, i.e., compared to Fig. 2 the quality of the deconvolution has increased threefold. As the comparison shows, the sharp Fermi edge is the most impor-

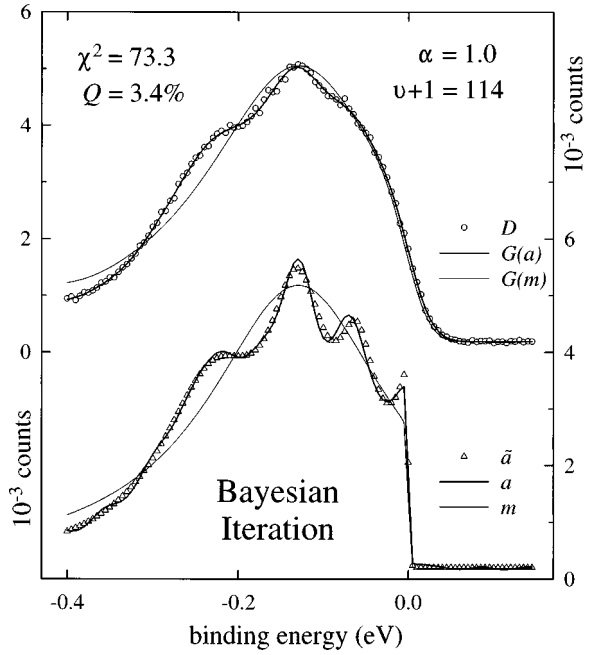

FIG. 4. Iterative Bayesian deconvolution a and $G(a)$ where the model is a Lorentzian multiplied by the Fermi-Dirac function plus a small dark count. The number of iterations $\nu+1=114$ is adjusted to obtain $\chi^2\approx 73$ as in Figs. 2 and 3.

tant spectral feature that should be incorporated into the model. It is also beneficial that the model of Fig. 3 does not contain the noise inherent in the $m_u=a^{(0)}=D$ case of Fig. 2 since a now interpolates smoothly between neighboring E_j. In Fig. 4 we replace the constant spectral function by one Lorentzian, truncated again by the Fermi-Dirac function, determining its intensity, energy position, and width and also the dark count by the usual least-squares fit procedure between D and $G(m)$, i.e., getting $\chi^2(m)$ of Eq. (4) to a minimum. We always perform the least-squares fit using the efficient Marquardt-Levenberg algorithm.[8] As seen in Fig. 4, the convolution of this model reproduces at least the gross features of D and has thus only a small negative bias. The quality of the iterative deconvolution, however, is barely higher than that of Fig. 3, showing that the Bayesian iteration is insensitive to the choice of the starting model as long as it is smooth, except for structures such as the Fermi edge, which is much narrower than the spectrometer function: Such structures should be incorporated if they are known.

In Figs. 2–4 the number of iterations is adjusted to obtain $\chi^2\approx 73$ to facilitate the comparison with the Bayesian analysis to be presented below. The minimum in Q actually occurs at slightly higher χ^2 values, but the difference to the a shown in Figs. 2–4 is hard to discern. The numbers for the model of Fig. 4 are $Q_{\min}=2.9\%$ at $\chi^2=75.0$. This is still at the low end of the $F(\chi^2)$ distribution (3) which has its maximum at $\chi^2=n-2=108$, i.e., $F(75)/F_{\max}=0.04$: For the best agreement between $a^{(\nu)}$ and $\tilde{a}$ it is apparently advantageous to allow some of the accidental structure caused by the noise in the final $a^{(\nu)}$ and $G^{(\nu)}$; however, $\chi^2\ll n-2$ should of course be avoided since this would produce a large $a^{\rm osc}$ component of $a^{(\nu)}$. Since we know neither $\tilde{a}$ nor Q for a true experimental spectrum, it is important to realize that there always is a continuum of possible deconvolutions a with different probabilities: The a obtained from the Bayesian iteration corresponding to χ^2 between the maximum and the low end of the $F(\chi^2)$ distribution all have relative high probabilities and the most probable $a=\hat{a}$ is likely to fall into this range. The reader might wonder why we do not just give $\hat{a}$ instead of this rule of thumb, but since finding $\hat{a}$ is equivalent to reinventing the Bayesian analysis, we move on to the next section.

III. BAYESIAN ANALYSIS

As already mentioned, a spectral function a with a large $a^{\rm osc}$ component is highly unlikely to conform with any reasonable physical model m of a even if it fulfills the $\chi^2\approx n$ test. We now investigate the crucial role of the model m in selecting physically acceptable a. We first sketch the important results presented in detail by Skilling,[2] starting by defining the conditional probability distribution $p(D|a)$ for the data D, given the spectral function a as the fixed condition,

$$dp_D=p(D|a)d^nD,$$

$$p(D|a)=\prod_k (2\pi D_k)^{-1/2}\exp(-\chi^2/2), \tag{9}$$

where dp_D is the probability that $G(a)$ falls into the range $d^nD=\Pi_k dD_k$ centered at D and χ^2 is defined in Eq. (4). The probability distribution $p(D|a)$ is approximately normalized for $\sqrt{D_k}\gg 1$, i.e., $\int dp_D\approx 1$. The least-squares fit used earlier corresponds to $p(D|a)\rightarrow\max$. Similarly, the conditional probability distribution $p(a|m,\alpha)$ of the spectral function a, given the model m and the regularization parameter α as the fixed conditions, is defined by

$$dp_a=p(a|m,\alpha)d^na,$$

$$p(a|m,\alpha)=\prod_i (2\pi a_i/\alpha)^{-1/2}\exp(\alpha S), \tag{10}$$

with dp_a the probability that m falls into the range $d^na=\Pi_i da_i$ centered at a and $\int dp_a\approx 1$ again for $\sqrt{\alpha a_i}\gg 1$. The entropy S is given by

$$S=\sum [a_i-m_i-a_i\ln(a_i/m_i)]$$

$$\approx-\sum [(a_i-m_i)^2/a_i]/2=-\chi_a^2/2, \tag{11}$$

where the right-hand side is valid for $a_i\approx m_i$ only, in which case Eq. (10) shows a strong similarity to Eq. (9), except for the regularization parameter α, which transforms a and m into fictitious counts by $a\rightarrow\alpha a$ and $m\rightarrow\alpha m$ when Eqs. (10) and (11) are derived from the Poisson distribution.[2] In fact, $\alpha\chi_a^2$ is obtained from χ^2 of Eq. (4) using $D\rightarrow\alpha a$ and $G\rightarrow\alpha m$: The distribution (9) and the approximate form of Eq. (10) valid for $a_i\approx m_i$ are completely equivalent in these ''counts.''

The basic problem encountered in Sec. II is to weed out the unphysical deconvolutions a containing a large $a^{\rm osc}$ component. To solve this problem we first join Eqs. (9) and (10) using the Bayes theorem[2,7] for conditional probabilities:

$$p(D,a|m,\alpha)=p(D|a,m,\alpha)p(a|m,\alpha)$$
$$=p(a|D,m,\alpha)p(D|m,\alpha). \quad (12)$$

The two distributions (9) and (10) are statistically independent, i.e., the joint distribution (12) is just the product of these two distributions, or $p(D|a,m,\alpha)=p(D|a)$ and thus

$$p(D,a|m,\alpha)\propto\exp(\beta), \quad \beta=-\chi^2/2+\alpha S, \quad (13)$$

where $p(D,a|m,\alpha)d^nDd^na$ is the probability that $G(a)$ falls into d^nD at D and simultaneously m into d^na at a, given m and α. This joint distribution gets large if both $G(a)$ is close to D and m is close to a, with the relative importance of the respective closeness governed by α. Any reasonable physical model m must not contain the spurious oscillations $a^{\rm osc}$ since they, by their definition (6), do not noticeably influence the closeness between D and $G(m)$ or $G(a)$. Thus $p(D,a|m,\alpha)$ can be large only if the $a^{\rm osc}$ component of a is also small since there would be large deviations between a and m otherwise. We thus obtain the most probable spectral function $\hat{a}$ with the smallest possible $a^{\rm osc}$ by requiring

$$p(D,a|m,\alpha)\propto\exp(\beta)\rightarrow\max \quad \text{or} \quad \partial\beta/\partial a_i=0, \quad (14)$$

leading to the system of n nonlinear equations for the n unknowns $\hat{a}_i$,

$$\hat{a}_i=m_i\exp(\Delta_i/\alpha). \quad (15)$$

Here Δ_i is given by Eq. (7) of Sec. II, with $G_k^{(\nu)}$ replaced by $G_k(\hat{a})$, and there is also a close connection between Eqs. (5) and (15): The only difference is that using $m_i=a_i^{(\nu)}$ we allowed for an iterative improvement of the model in Sec. II while the model is constant in Eq. (15), requiring $a_i^{(\nu+1)}\rightarrow\hat{a}_i$ here. Applying the general theorem of Bayes (12) to the deconvolution problem is thus equivalent to equating Δ_i in Eq. (15) with the weighted average (7) of the relative deviations between D and $G(\hat{a})$. This is a useful way to look at the problem, which even helps us to dodge a severe technical difficulty encountered in solving Eq. (15). We discuss this obstacle and our method to remove it in the following two paragraphs, completing the fundamentals of the Bayesian analysis afterward.

Usually the n nonlinear equations (15) are solved for the $\hat{a}_j$ by iteratively solving the corresponding set of n linear equations[9] obtained in this case by a linear expansion of the $\ln(\hat{a}_i/m_i)$ term. However, considerable convergence problems are encountered for $\alpha\ll 1$ typical for a low-quality model deviating markedly from $\hat{a}$ and/or for a high signal-to-noise ratio. The reason for these difficulties is that, e.g., at step ν of the iteration, $a_j^{(\nu)}<0$ may occur for some final solutions $\hat{a}_j$ that are close to zero, in spite of the fact that $\hat{a}_j>0$ is required by Eq. (15) for all j.

We simply bypass these convergence problems by solving the nonlinear equations (15) iteratively for the Δ_j instead: Since $\Delta_i<0$ is *required* by Eq. (15) for $\hat{a}_i<m_i$ and $\Delta_i>0$ for $\hat{a}_i>m_i$, respectively, the solutions of the corresponding system of n linearized equations

$$\sum_{j=1}^{n} M_{ij}\Delta_j^{(\nu)}=C_i \quad (16)$$

at the step ν of the iteration quickly converge to the solutions Δ_j of Eq. (15). The coefficients of Eq. (16) are

$$M_{ij}=H_{ij}\frac{m_j}{\alpha}\exp(\Delta_j^{(\nu-1)}/\alpha)+\delta_{ij}, \quad H_{ij}=\sum_{k=1}^{n}\frac{G_{ik}G_{jk}}{D_k}=H_{ji}; \quad (17)$$

$$C_i=\sum_{k=1}^{n}[G_{ik}(1+B_k/D_k)], \quad B_k=\sum_{h=1}^{n}m_h(\Delta_h^{(\nu-1)}/\alpha-1)\exp(\Delta_h^{(\nu-1)}/\alpha)G_{hk}. \quad (18)$$

The iteration is best started with $\Delta_j^{(0)}=0$ for all j corresponding to $a_j^{(0)}=m_j$ according to Eq. (15). After convergence is obtained, the $\hat{a}_i$ are calculated by inserting the Δ_i in Eq. (15). We also verified that the old method of solving Eq. (15) for $\hat{a}_j$ and the method presented here lead exactly to the same numerical values of $\hat{a}_j$. This comparison is of course possible only as long as convergence is achieved with the old method.

We now return to the basics of the Bayesian analysis. While we can solve Eq. (15) for any given m and α, we still do not know the optimal α_0. This problem is attacked by the regularization procedure: Instead of focusing on the most probable spectral function $\hat{a}$, Skilling[2] proposed to take the continuum of all possible a into account, i.e., to determine the maximum of the joint probability distribution (13), integrated over all a:

$$\alpha\rightarrow\alpha_0 \quad \text{if} \quad p(D|m,\alpha)=\int d^na\ p(D,a|m,\alpha)\rightarrow\max. \quad (19)$$

In Eq. (19) the Bayes theorem (12) and $\int d^na\ p(a|D,m,\alpha)=1$ have been used. The integration can be carried out by expanding $\beta(a)$ of Eq. (13) around the most probable $\hat{a}$, where Eq. (14) ensures that no linear terms show up, i.e.,

$$\beta(a)\approx\beta(\hat{a})+\frac{1}{2}\sum_{i,j}\frac{\partial^2\beta(\hat{a})}{\partial a_i\partial a_j}(a_i-\hat{a}_i)(a_j-\hat{a}_j), \quad (20)$$

and transforming Eq. (20) to its purely quadratic form. The result after the regularization (19) is

$$p(D|m,\alpha_0)=\frac{\Pi_k(2\pi D_k)^{-1/2}\exp[\beta(\hat{a})]}{\sqrt{\det(I+\alpha_0^{-1}[\hat{a}]H)}}, \quad (21)$$

where I is the identity matrix, $[\hat{a}]$ is the diagonal matrix, and H is defined in Eq. (17). Equation (21) is identical to the corresponding expression given by Skilling[2] if $R_{jk}=R_{kj}$, which is the case for our choice of R_{jk} shown in Eq. (2).

The regularization culminating in Eq. (21) completes what we call the Bayesian analysis. We now show that this analysis goes beyond the least-squares fit on three counts. The numerator of Eq. (21) is proportional to the maximum of the joint distribution (13) for the most probable $\hat{a}=\hat{a}(\alpha_0)$. Thus, in addition to the quadratic deviations $\chi^2(\hat{a})$ between the data D and the convolution of $\hat{a}$, the Bayesian analysis

also accounts for the corresponding deviations $\alpha_0 S$ between $\hat{a}$ and the model m. Furthermore, the denominator of Eq. (21) takes care of the fact that there is a continuum of possible spectral functions around $\hat{a}$: As shown below, the width of this continuum increases from zero, α_0 decreases from infinity, and $\alpha_0 S$ decreases from zero as m starts to deviate from the true spectral function $\tilde{a}$, causing a corresponding increase in the denominator and thus a further decrease of Eq. (21) itself, in addition to the decrease caused by the $\alpha_0 S$ exponent in the enumerator. In comparison, the least-squares fit determines the minimum of $\chi^2(m) \neq \chi^2(\hat{a})$ only.

The most important application of the Bayesian analysis is the quantitative comparison of different theoretical models m in terms of the measured data D by calculating the ratio of the corresponding probabilities (21). The existence of such a quantitative comparison provides of course the ultimate motivation of taking the spectra in the first place. Any other data analysis, including the least-squares fit, allows a qualitative comparison only since (i) the difference between $\chi^2(m)$ and $\chi^2(\hat{a})$, (ii) the deviations between m and $\hat{a}$ described by $\alpha_0 S$, and also (iii) the continuum of the possible spectral functions around $\hat{a}$ are not accounted for.

A. Role of the model

In the Bayesian analysis the role played by the model m in determining the quality of the most probable spectral function $\hat{a}$ is not readily apparent. We therefore consider the limiting case $\hat{a}_i \approx m_i$, i.e., $|\Delta_i/\alpha_0| \ll 1$ using Eq. (15), and assume that the regularization has been carried out, i.e., that α_0 and $\hat{a}$ are known for the given model m and the measured spectrum D. The nonlinear equations (15) can then be linearized as $\hat{a}_i - m_i \approx m_i \Delta_i/\alpha_0$. Rewriting, summing, and taking Eq. (11) into account, we obtain

$$\chi_a^2 \approx \sum (\hat{a}_i - m_i)^2/m_i \approx \sum (\tilde{a}_i - m_i)^2/m_i \approx \alpha_0^{-2} \sum m_i \Delta_i^2, \tag{22}$$

where we replace the most probable $\hat{a}_i$ by the true $\tilde{a}_i$, which should be a very good approximation since χ_a^2 does not measure the individual but only the average squared deviations between $\hat{a}_i$ and m_i. Thus Eq. (22) leads to

$$\alpha_0 \approx \sqrt{\sum m_i \Delta_i^2} \Big/ \sqrt{\sum (\tilde{a}_i - m_i)^2/m_i} \tag{23}$$

and also, using Eqs. (11) and (23), to

$$\alpha_0 S(\tilde{a}) \approx -\tfrac{1}{2} \sqrt{\sum m_i \Delta_i^2} \sqrt{\sum (\tilde{a}_i - m_i)^2/m_i}, \tag{24}$$

which, for $m_i \to \tilde{a}_i$, show that $\alpha_0 \to \infty$ while $\alpha_0 S(\tilde{a}) \to 0$. We notice that the noise contained in D requires $\Sigma m_i \Delta_i^2 > 0$ even in this limit. The $\alpha_0 S(\tilde{a}) \to 0$ limit is apparently the reason that the Bayesian analysis is often referred to as the maximum-entropy method, which is actually a misnomer since the joint probability distribution of Eq. (13) reaches a maximum for the maximum value of the exponent $\beta = -\chi^2/2 + \alpha_0 S$, not just of the entropy S alone.

Introducing α_0 of Eq. (23), $\alpha_0 S(\tilde{a})$ of Eq. (24), and β of Eq. (13) into Eq. (21), $p(D|m,\alpha_0)$ is seen to decrease as m moves away from $\tilde{a}$. We can even derive an approximate expression of Eq. (21) showing this dependence on m explicitly, valid for $m_i \approx \tilde{a}_i$ corresponding to $\alpha_0 \gg 1$: Using the right-hand side of Eq. (11), the distribution (10) is a product of normalized Gaussians centered at m_i, with the small width parameter $\sigma_{ai} = (a_i/\alpha_0)^{1/2}$. We approximate this product by a product of δ functions and introduce it into Eq. (19), taking Eq. (13) into account. The integration (19) then results in

$$p(D|m,\alpha_0 \gg 1) \approx \prod_k (2\pi D_k)^{-1/2} \exp[-\chi^2(m)/2]. \tag{25}$$

For high-quality models, the Bayesian analysis including the regularization condition (19) reduces to the usual least-squares fit between $G(m)$ and D.

In the limit of $m = \tilde{a}$ for which $\alpha_0 \to \infty$, the basic system of nonlinear equations (15) yields $\hat{a} = m$ and thus $\hat{a} = \tilde{a}$. Since in general $\hat{a} = \tilde{a} + a^{\mathrm{osc}}$ according to Eq. (6), it follows that $a^{\mathrm{osc}} = 0$ in this case. As m starts to deviate from $\tilde{a}$, we obtain the decrease of $p(D|m,\alpha_0)$ discussed above and the decrease of α_0 described by Eq. (23). Furthermore, since the coefficients m_i and α_0 of the nonlinear system of equations (15) are changed with respect to the $m = \tilde{a}$ case, $\hat{a} \neq \tilde{a}$ or $a^{\mathrm{osc}} \neq 0$: A decrease in the quality of m causes a concurrent increase in the amplitude of a^{osc}. Thus a model with a small value of $p(D|m,\alpha_0)$ is unlikely to conform with D since it generates the large oscillatory deviation a^{osc} between $\hat{a}$ and $\tilde{a}$. This statement is of course valid in spite of the fact that $\tilde{a}$ is not known in general. The most important consequence of this behavior is the possibility to evaluate different theoretical models m in terms of the measured spectrum D according to their respective probability distribution $p(D|m,\alpha_0)$ quantitatively.

We obtain some further insight from $p(a|D,m,\alpha_0)$, which is the normalized probability distribution of a, given D, m, and α_0. Since the joint distribution $p(D,a|m,\alpha_0)$ reaches its maximum at $a = \hat{a}$ and the integrated joint distribution $p(D|m,\alpha_0)$ no longer contains a, the Bayes theorem (12) requires a maximum value of $p(a|D,m,\alpha_0)$ also at $a = \hat{a}$, given by

$$p(\hat{a}|D,m,\alpha_0) = \sqrt{\prod_i (2\pi \hat{a}_i)^{-1} \alpha_0^n \det(I + \alpha_0^{-1}[\hat{a}]H)}, \tag{26}$$

which clearly goes to infinity as $\alpha_0 \to \infty$ or $m = \tilde{a}$, i.e., $p(a|D,m,\alpha_0)$ is a product of δ functions in this limit. As m starts to deviate from $\tilde{a}$, the width of $p(a|D,m,\alpha_0)$ increases, accommodating a broader range of acceptable a around $\hat{a}$, while $p(\hat{a}|D,m,\alpha_0)$ and $p(D|m,\alpha_0)$ both decrease simultaneously. The most important quantity, however, is the integrated joint probability distribution (21), which we normalize to its most probable maximum value by

$$P = \frac{p(D|m,\alpha_0)}{p(D|m=\tilde{a},\alpha_0 \to \infty)} = \frac{\exp[\beta(\hat{a}) + (n-2)/2]}{\sqrt{\det(I + \alpha_0^{-1}[\hat{a}]H)}}, \tag{27}$$

obtained by using $\chi^2(\tilde{a}) = n - 2$ in Eq. (25), where $n-2$ corresponds to the maximum of the $F(\chi^2)$ distribution (3). It is important to note that P is determined by m and D *includ-*

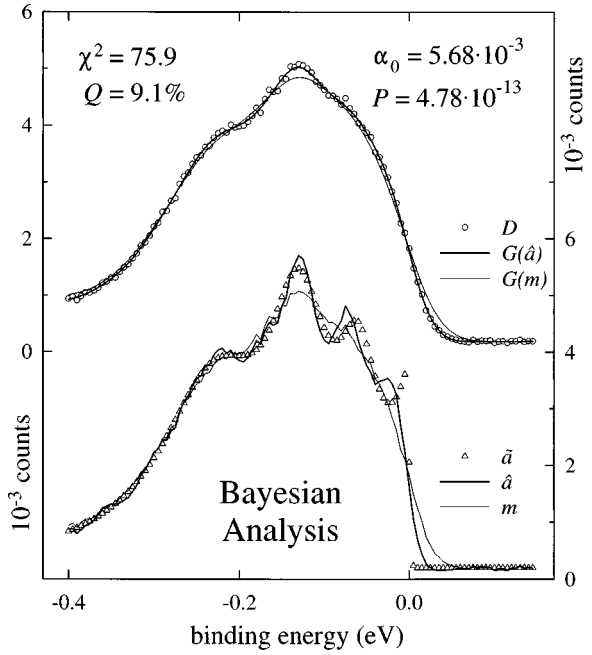

FIG. 5. The most probable spectral function $\hat{a}$ and $G(\hat{a})$, obtained by the Bayesian analysis. As in Fig. 2, the experimental spectrum D itself is taken as the model m.

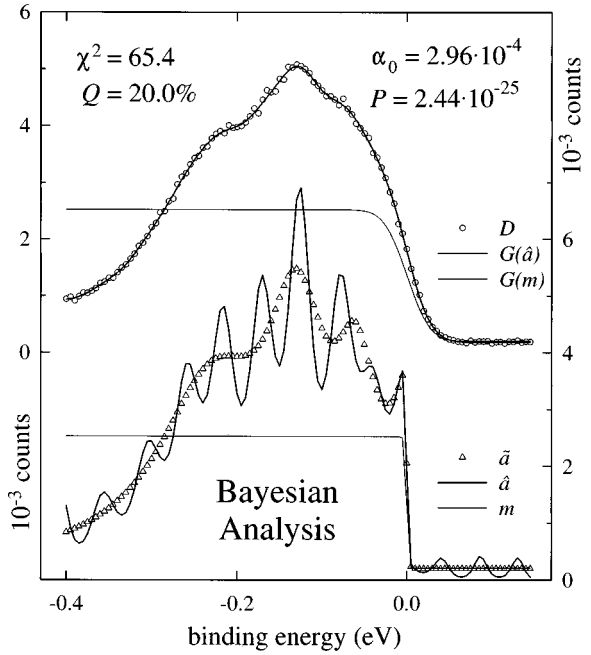

FIG. 6. The most probable spectral function $\hat{a}$ and $G(\hat{a})$, obtained by the Bayesian analysis. The model is the same as in Fig. 3.

ing its noise component: For fixed m, an increase in the signal-to-noise ratio causes a concurrent increase in $\chi^2(m)$ and a *decrease* in P; since more details have been uncovered in D, the old P value can only be restored by a more detailed m.

We intentionally avoid to attach the usual names (prior, likelihood, posterior, evidence) to the distribution functions in Eq. (12) since we want to show that one does not really need them, at least in the present case. They might actually be misleading since there is no temporal order in the problem: The measured spectra D are there, waiting to be analyzed, and since the model m is the given condition, i.e., standing to the right of the vertical bar in Eq. (27), we are free to use any model, as long as we are not withholding knowledge, which would be very un-Bayesian indeed. What we know for sure is that D, by the convolution, is a broadened version of the true spectral function $\tilde{a}$ to which noise has been added, and although we do not know $\tilde{a}$, we do know that the model should come reasonably close to it, i.e., we should incorporate at least the gross features of D into $G(m)$. For example, if D shows a pronounced structure, it is certainly not a good idea to use $m=\text{const}$ and the penalty will be low values for P and α_0 and a large a^{osc} component in $\hat{a}=\tilde{a}+a^{\text{osc}}\neq m$, *not* an approximate reproduction of the bad model as usually assumed without proof, and this is true for *any* bad model, not only for $m=\text{const}$: The Bayesian analysis is incorruptable and yields $\hat{a}\approx m\approx\tilde{a}$ together with $|a^{\text{osc}}/\tilde{a}|\approx 0$ for the largest values of $P\approx 1$ and $\alpha_0\gg 1$ only.

In the ideal case the functional form of the theoretical model m is known and D is distorted by the spectrometer function R_{jk} and the noise only. We then obtain $P\approx 1$ and $\hat{a}\approx m\approx\tilde{a}$ by just optimizing the parameters of the model. Usually, however, D is corrupted in addition by a plethora of effects outside of the hopefully dominating theoretical model to be investigated. In spite of these odds, the Bayesian analysis yields reliable results for two competing theoretical models m_1 and m_2, provided Ockham's razor[10] is applied rigorously, which simply requires one to use as few parameters as possible under the circumstances: Although we obtain $P_{m1}\ll 1$ and $P_{m2}\ll 1$ since the small set of parameters does not adequately describe these outside effects, their influence will essentially cancel in the P_{m1}/P_{m2} ratio. A fair contest requires of course to use the same number of parameters in both models. As a consequence of the small P_m values we obtain a noticeable a^{osc} component in $\hat{a}$, but since we admit only models with $P_m\geqslant P_{mu}$, where $m_u=D$ is the only unbiased model independent of any theoretical assumptions, $|a^{\text{osc}}/\tilde{a}|\ll 1$ is still ensured.

B. Computer simulation of the analysis

Performing computer experiments as before, $\tilde{a}$ is known and the conclusions given above can be checked. In the following figures we now include the values for α_0 and P. Figure 5 gives the result of the Bayesian analysis for $m_u=D$. The result is nearly identical to that of Fig. 2, where we had carried the iterative Bayesian deconvolution to roughly the same χ^2. In comparison, the advantage of the Bayesian analysis is that it is independent of the scientist performing it, requiring no judgment about the number of iterations as in the case of the iterative Bayesian deconvolution.

Since the $m=\text{const}/[\exp(E/kT)+1]+C_0$ model (where C_0 denotes dark count) contains the additional information about the Fermi edge, we obtain the excellent agreement between a and $\tilde{a}$ in the case of the iterative Bayesian deconvolution shown in Fig. 3. In contrast, the corresponding Bayesian analysis leads to the disaster evident in Fig. 6 because this model has a strong negative bias for $E<0$ and, as opposed to Fig. 3, the Bayesian analysis does not allow for the iterative correction of m. Compared to the unbiased

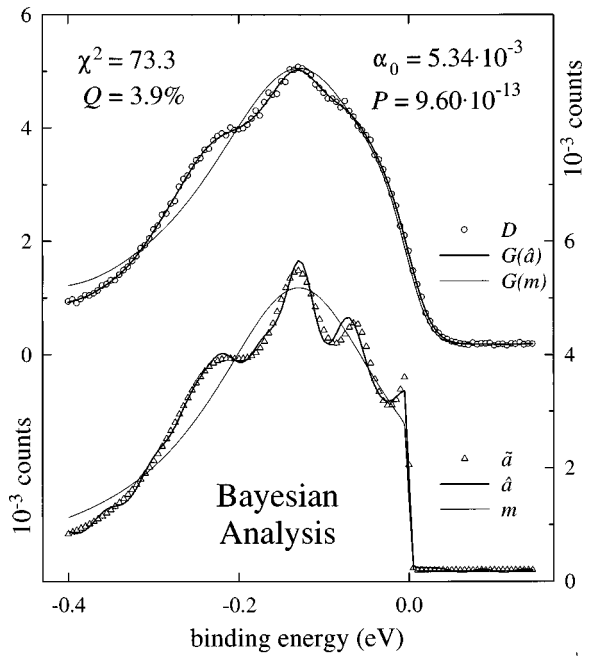

FIG. 7. The most probable spectral function $\hat{a}$ and $G(\hat{a})$, obtained by the Bayesian analysis. The model is the same as in Fig. 4.

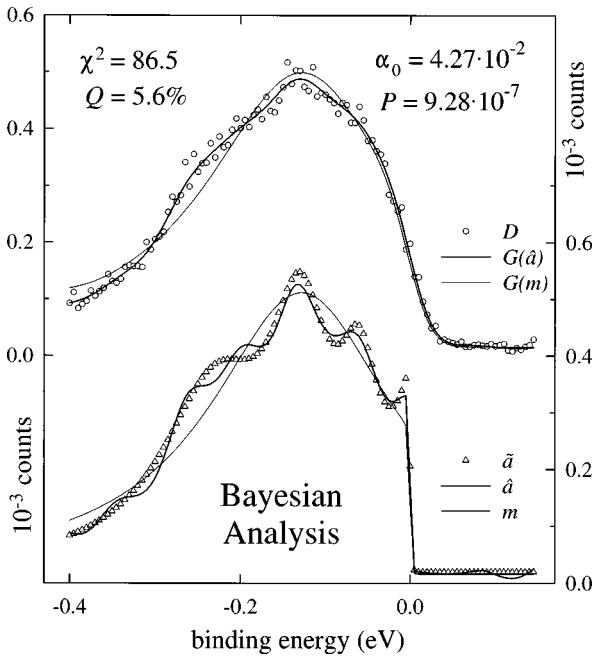

FIG. 8. The most probable spectral function $\hat{a}$, obtained by the Bayesian analysis as in Fig. 7, except for a 0.1 reduction of the counts in the experimental spectrum D.

$m_u = D$ case of Fig. 5, the value for P is 12 orders of magnitude smaller, corresponding to the large a^{osc} component evident in $\hat{a}$, and Q shows a pronounced increase. This is just what we anticipated from analyzing Eqs. (23)–(25). Characteristically, $G(\hat{a})$ is even closer to D than before since we obtain a lower χ^2, in spite of the disastrous failure of this model. Thus, in contrast to the iterative Bayesian deconvolution, a model m with a strong negative bias, i.e., large deviations between $G(m)$ and D, should never be used to perform the Bayesian analysis; any conclusions drawn from the $\hat{a}$ thus obtained are not trustworthy, in spite of $G(\hat{a}) \approx D$. As a rule of thumb, any acceptable model must produce a P at least as high as that of the unbiased model $m_u = D$.

A nice supporting example has been given by von der Linden.[4] He performed a Bayesian analysis of a spectrum D produced by Ar^+ ions scattered at a clean Ru(001) surface, using at first a small constant as the model, which we would classify as strongly negatively biased since D shows a pronounced single peak at the expected energy. The $\hat{a}$ contains a large a^{osc} component, which he calls ''ringing.'' He then tried a model composed of a constant and a Gaussian that takes the evidence in D into account. This model drastically reduces the ringing and, as we know from Eqs. (23)–(25), brings $\hat{a}$ much closer to $\tilde{a}$ simultaneously.

Returning to the computer experiment presented in Fig. 4, we recall that the $m = L/[\exp(E/kT)+1]+C_0$ model (where L denotes Lorentzian) contains four parameters that we optimize by a least-squares fit prior to the iterative Bayesian deconvolution. These four parameters could in principle also be determined by the Bayesian analysis, but this would be an extremely tedious job indeed, resulting in a model m as close as possible to the true spectral function $\tilde{a}$ and simultaneously $G(m)$ as close as possible to D, i.e., $\chi^2(m) \rightarrow \min$: The two procedures are equivalent, but the least-squares fit is faster and easier by a huge margin. As an example, Fig. 7 shows the result of the Bayesian analysis with the above model optimized already. We obtain $\chi^2(\hat{a}) = 73.3$, which is also the χ^2 value reached in Fig. 4 after $\nu+1 = 114$ iterations. As opposed to the huge differences found between Figs. 3 and 6, a in Fig. 4 and $\hat{a}$ in Fig. 7 are nearly identical since m has now only a small negative bias for $E<0$. Compared to the $m_u = D$ case of Fig. 5, P has doubled, with a corresponding decrease in Q: For the present model the positive bias of the known Fermi edge carries more weight than the slightly more negative bias in the $E<0$ region.

It is noteworthy that the a^{osc} components in the three Bayesian iterations of Figs. 2–4 and also in the Bayesian analysis presented in Figs. 5 and 7 are nearly identical. Even in Fig. 6, the maxima and minima of a^{osc} occur at about the same energies, although with a much larger amplitude. Actually, a close inspection shows that a^{osc} just follows the smoothed noise spectrum with an average period of about 0.05 eV, which is the FWHM of the spectrometer function. This is to be expected from Eq. (7), where the noise is averaged roughly over this width: Using Eq. (6) we obtain $G_k \approx G_k(\tilde{a})$ in Eq. (7) both in the final steps of the Bayesian iteration and in the Bayesian analysis, which shows that Δ_i is then roughly the convolution of the relative noise spectrum. Repeating the Bayesian analysis of Fig. 7 with the same true spectral function $\tilde{a}$ but using different Poisson distributed noise spectra in Eq. (1), we find correspondingly different a^{osc} components, along with variations in χ^2, Q, and P, as expected.

The result of the Bayesian deconvolution or analysis depends of course strongly on the signal-to-noise ratio. Figure 8 gives an example: The analysis has been carried out as in Fig. 7, except for reducing the number of counts to $\frac{1}{10}$, corresponding to a $\sqrt{10}$-fold increase in the relative amplitude of the noise. The higher value of α_0 shows that emphasis has been shifted from χ^2 towards the entropy S. The increase in

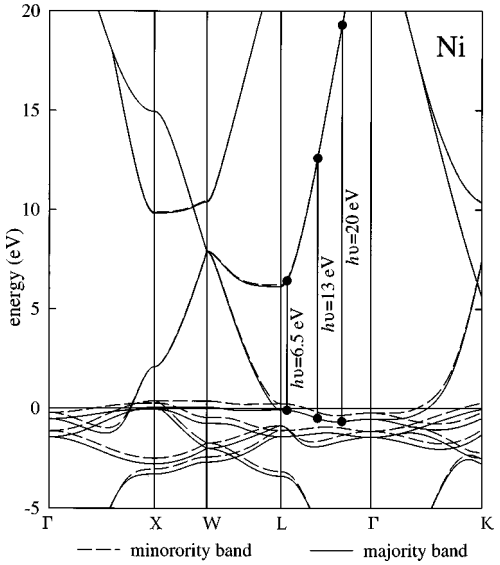

FIG. 9. Experimental band structure of Ni. Three typical direct transitions from the majority $3d$ band just below the Fermi energy $E_F=0$ eV that are observed in the photoelectron spectra for normal emission from (111) are indicated.

P indicates that $G(m)$ is more likely to fit D, given the increase in the relative bandwidth of the noise. Since it is difficult now to distinguish between a^{osc} and the genuine structure in $\hat{a}$ (we pretend not to know $\tilde{a}$), improvement of the model is mandatory. Fortunately, the improvement is also possible since both the raw data D and $G(\hat{a})$ show that at least two peaks are involved. Reanalyzing with the improved model containing two Lorentzians will increase P and simultaneously reduce the amplitude of a^{osc}, identifying it as spurious and the remaining common structure as genuine, quite analogously to going from the bad model of Fig. 6 to the improved model of Fig. 7. Thus a decrease in the signal-to-noise ratio requires an increase in the quality of the model, which is getting more difficult to achieve at the same time: For a sufficiently low signal-to-noise ratio we obtain $\alpha_0 \rightarrow \infty$ for both $m=L/[\exp(E/kT)+1]+C_0$ and for $m=\tilde{a}$, i.e., the quality of the spectrum D is so low that it is no longer possible to distinguish between a model containing one Lorentzian and the true spectral function with four Lorentzians.

IV. DO THE 3*d* ELECTRONS IN Ni FORM A FERMI LIQUID?

The experimental energy bands of Ni (Ref. 11) are reproduced in Fig. 9. They have been obtained by adjusting the parameters of the combined interpolation scheme for best agreement with the experimental angle resolved photoelectron spectra and the measured magneton number. Three typical transitions from the majority band that are observed in the photoelectron spectra for normal emission from (111) are indicated. The corresponding spectra obtained by us are presented in Fig. 10. They show both the majority and the minority spin peak of Ni, observed in *s* polarization for ten different photon energies, again under normal emission from the (111) surface. The polarization vector is parallel to [110], the angle of incidence is 22.5°, and the sample is kept at

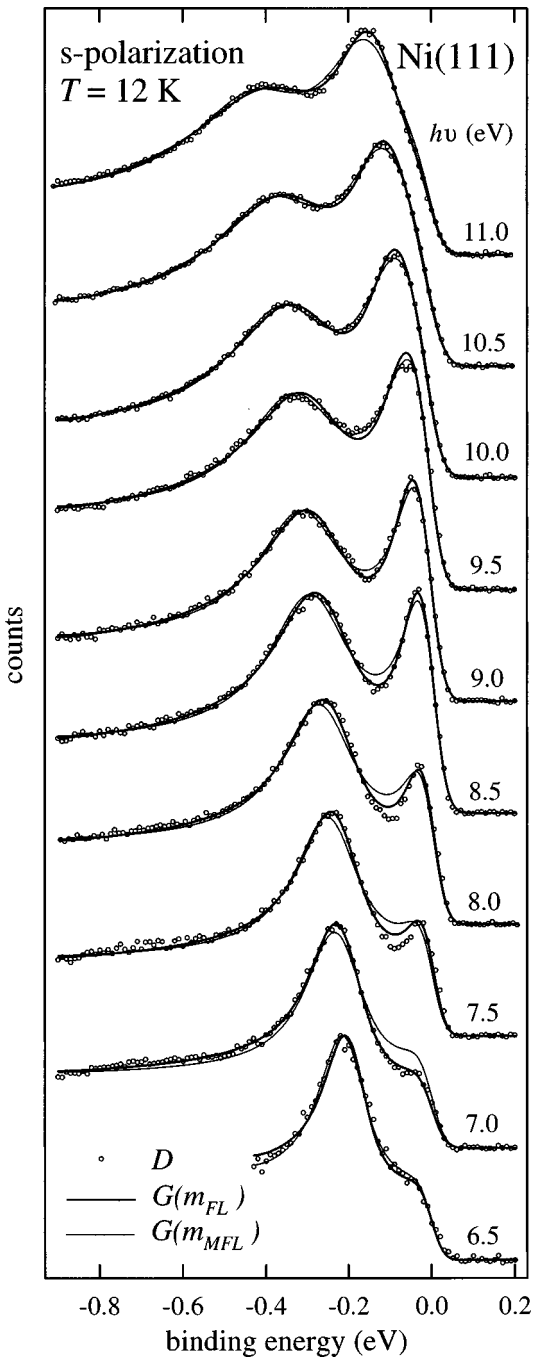

FIG. 10. Ten photoelectron spectra, obtained at $T=12$ K from Ni(111) in normal emission, for photon energies between $h\nu=6.5$ and 11 eV and *s* polarization (open circles). These spectra are all displayed with the same difference between the maximum count and the constant count above E_F and this normalization sets the scale for the convolution of the optimized FL and MFL models shown as heavy and light solid lines here and also for the deconvolutions shown in Figs. 11 and 12.

about 12 K. The measurements are performed at the 2-m Seya beam line of the Berliner Elektronenspeicherring für Synchrotronstrahlung. The electron analyzer, which allows simultaneous detection of eight coplanar directions, consists of four concentric spherical grids, forming a preretarding stage and a high pass filter, and of two grids that are part of two confocal ellipsoids of revolution, forming an electron mirror, i.e., a low pass filter. One focus of these ellipsoids of

revolution coincides with the center of the spherical grids. We determine the combined energy resolution of this electron analyzer and the 2-m Seya-Namioka monochromator from the 12-K Fermi edge of a polycrystalline Pt sample. The FWHM of the spectrometer function R_{jk} in Eq. (1) is 50 meV at $h\nu=6.5$ eV and 60 meV at $h\nu=11$ eV, respectively, and the angle resolution is 2° FWHM.

The electropolished surface of the sample is cleaned *in situ* by repeated cycles of 500-eV Ar ion bombardment at 600 K and annealing at 850 K. After this cleaning procedure we observe sharp low-energy electron diffraction spots. Immediately after cleaning, the sample is brought in thermal contact with the cryostat and cooled to about 12 K. The high annealing temperature gives rise to the relatively high base pressure of 1×10^{-10} mbar even after cooling down, which, together with the high reactivity of the Ni surface, causes adsorption of residual gases: The cleaning procedure is repeated after 2 h since a noticeable increase in the background of secondary electrons can be detected by then.

A. Two basic models

As the peaks in Fig. 10 disperse away from E_F, their width increases markedly. This is the behavior that we expect qualitatively if the $3d$ electrons in Ni form a Fermi liquid (FL). Actually, Claessen *et al.*[12] have shown that the quasiparticle spectral function

$$a(E,k)=\pi^{-1}\mathrm{Im}\,\Sigma/[(E-E_B)^2+(\mathrm{Im}\Sigma)^2], \tag{28}$$

with the FL form of the imaginary part of the self-energy Σ,

$$\mathrm{Im}\,\Sigma_2=\beta_2E^2,\quad E\approx E_F=0, \tag{29}$$

valid for $T\to0$ and for E close to E_F, can be used to describe the measured angle resolved photoelectron spectra of the two-dimensional metal 1-T-$TiTe_2$, albeit only extremely close to E_F. This is actually a favorable example since the two-dimensional character of 1-T-$TiTe_2$ prevents the finite escape depth of the photoelectron to contribute to the observed line shape. In the case of the three-dimensional $E(k)$ of Ni, the additional broadening caused by the finite escape depth has to be incorporated. If we had complex initial and final bands E_i and E_f with Lorentzian width Γ_i and Γ_f, respectively, the total width Γ would be[13]

$$\Gamma=\frac{\Gamma_i+|r|\Gamma_f}{1-r},\quad r=\frac{\partial E_i}{\partial k_\perp}\Big/\frac{\partial E_f}{\partial k_\perp}, \tag{30}$$

where r is the ratio of the slopes in the respective bands. We actually select Ni(111) for this investigation since, as is evident from Fig. 9, the slope of the initial band E_i close to the top of the $3d$ bands is small compared to the slope of the nearly free electron final band E_f, resulting in $r\ll1$: The influence of the finite escape depth, represented by Γ_f, is strongly reduced and $1-r$ in the denominator of Eq. (30) can be replaced by 1. Taking Eqs. (28) and (30) as a guide, we assume that the model

$$m(E)=f(E)\sum_{s=\uparrow}^{\downarrow}C_s\Gamma_s/[(E-E_s)^2+\Gamma_s^2],$$
$$\Gamma_s(E)=\mathrm{Im}\,\Sigma(E)+|r|\Gamma_{fs} \tag{31}$$

is a good approximation of the true spectral function. In Eq. (31) the ratio r is defined as in Eq. (30) and the sum over the two spin components is multiplied by the Fermi-Dirac function $f(E)$.

If we would perform the Bayesian analysis, using the data of Fig. 10 and the model (31) with the FL form (29) of Im Σ, little would be gained, however, since the absolute value P of Eq. (27) does not allow us to calculate something like ''the probability for the validity of the Fermi liquid model'' on an absolute scale since such a quantity simply does not exist. Even a well tested theory can claim a higher probability only since, by comparison, it describes the experiments better than other competing theories. We thus need a competing model, but we also have to take Ockham's razor[10] into account: As shown above, we can increase P by improving the model, adding more parameters to it. We only can claim advantage for a model with a high P value if it contains the same (or a smaller) number of parameters as (than) the rival model with a lower P.

The marginal Fermi-liquid (MFL) model[6] has been introduced to explain the unusual properties of the high-temperature superconductors, attributed to the strong correlations between the conduction electrons. Since the $3d$ electrons in Ni are also strongly correlated, we decide to investigate this model in competition with the FL model. Instead of Eq. (29), the MFL model requires

$$\mathrm{Im}\,\Sigma_1=\beta_1|E|,\quad E\approx E_F=0. \tag{32}$$

Since the two models have the same number of parameters, a fair contest is ensured.

B. Refining the models

The energy dependence of the electron escape depth or mean free path Λ_{sm} for the elements, measured in monolayers (ML), is roughly given by the universal curve[14]

$$\Lambda_{sm}=A/(E_s+h\nu)^2+B\sqrt{a_0(E_s+h\nu)},$$
$$\rho=m/a_0^3, \tag{33}$$
$$\Lambda_s=a_0\Lambda_{sm},$$

where $A=538$ ML $(\mathrm{eV})^2$, $B=0.41$ ML $(\mathrm{nm\,eV})^{-1/2}$, ρ is the density, m is the atomic mass, and a_0 is the monolayer thickness. As in Eq. (31), the index s refers to the spin-up and spin-down components, respectively, E_s is the corresponding binding energy, and $h\nu$ is the photon energy. Since we define $E_F=0$ as the zero of the energy scale, $E_s+h\nu$ is the energy above E_F. If ξ is the imaginary part of the wave vector and ψ the wave function of the photoelectron in the crystal, the intensity I will decay according to $I=|\psi|^2\propto\exp(-2\xi z)=\exp(-z/\Lambda_s)$ and thus

$$\Gamma_{fs}=(\partial E_f/\partial k_\perp)\xi=(\partial E_f/\partial k_\perp)/2\Lambda_s, \tag{34}$$

i.e., the last term on the right-hand side of Eq. (31) can be written as

$$|r|\Gamma_{fs}=|\partial E_i/\partial k_\perp|/2\Lambda_s. \tag{35}$$

The slope of the initial band that is needed in Eq. (35) could be taken from the band structure[11] shown in Fig. 9, but since

our measurements are taken at 12 K and are better resolved than the ones used in Ref. 11, we perform a preliminary least-squares fit to our spectra, of which Fig. 10 presents a selection only, using a sum of two Lorentzians as the model, truncated again by the Fermi-Dirac function. We then use the binding energies of these Lorentzians to correct the parameters of the combined interpolation scheme and to recalculate the spin-up and spin-down components of the Λ_3 initial band at the top of the $3d$ bands, together with the slopes required in Eq. (35).

The experimental spectra also contain background electrons of different origin that have to be incorporated into the model. The constant count above E_F displayed by all experimental spectra in Fig. 10 is caused by stray light and amounts to about 15% of the maximum count. We determine this contribution for each $h\nu$ as the arithmetic mean of the last 15 data points at the right-hand side of the spectra and simply add it to the model (31), assuming that this stray light contribution remains constant also below E_F. The remaining background consists of quasielastically scattered electrons, of secondary electrons, and of electrons scattered in the analyzer after having surmounted the high pass filter. We discuss these three contributions in turn below.

The photoelectrons may be scattered quasielastically by surface irregularities resulting from incomplete recrystallization after the argon ion bombardment or from adsorbed atoms or molecules. The spectral distribution of this quasielastic background observed in the [111] direction is approximately proportional to the angle integrated spectrum of the elastically emitted photoelectrons, which is again close to the emission into the three {110} mirror planes containing the [111] axis:[15,16] The two exchange split peaks are shifted by up to about ± 0.25 eV with respect to their energy position at normal emission.[15] This causes a sizable signal even at energies where the contributions of both the elastic and the inelastically scattered electrons are low. An example of this situation is the Fermi edge of the $h\nu = 6.5$ eV spectrum, which is dominated by the quasielastically scattered electrons.

In the Shirley approximation,[17] the intensity of the secondary electrons at the energy E is proportional to the integral over the angle integrated spectrum above E, containing both the elastic and the inelastically scattered electrons. The electrons scattered in the electron analyzer can be described by a similar expression, except that the integration is now carried over all electrons above E emitted along [111] only.

A detailed description of these three background contributions would require at least three additional parameters in model. While this procedure will lead to a nearly perfect least-squares fit with $\chi^2 \approx n$, it will also effectively mask any deficiencies the FL or the MFL model might have, in clear violation of Ockham's razor. On the other hand, some crude model of the background is needed since we would obtain $\chi^2 \gg n$ otherwise. We decide to lump the three contributions together, describing them by the single constant C_u, which is added to the model (31) after multiplying it by the Fermi-Dirac function. This approximation might not be that crude after all since the contribution of the quasielastically scattered electrons is large close to E_F, where the intensity of the secondary electrons and the electrons scattered in the analyzer is small, while the situation is reversed at higher binding energies. Assuming that the escape depth is given by the universal curve (33), we end up with the six adjustable parameters $C_{\uparrow,\downarrow}$, $E_{\uparrow,\downarrow}$, β, and C_u.

C. Possible sources of error

The angle resolution will also contribute to the observed line shape. In the case of Ni close to the [111] axis, the surface of constant energy for the uppermost exchange split d band is approximately a plane perpendicular to [111] and the final band approximately nearly free electronlike. In this case $E_{\rm kin}(\Theta) - E_{\rm kin}(0) \approx |r| E_{\rm kin}(\Theta) \sin^2(\Theta)$, where $E_{\rm kin}$ is the kinetic energy of the photoelectron in vacuum and $|r| \ll 1$ is the ratio of the slopes in the initial and final bands defined in Eq. (30). The spread $\Delta\Theta$ in the polar angle Θ causes the energy spread $\Delta E \approx |r| E_{\rm kin}(\Theta) \sin(2\Theta) \Delta\Theta$. Since ΔE is minimal for $\Theta \to 0$ we investigate normal emission with $\Theta = 0$ from Ni(111) only. We determine ΔE caused by $\Delta\Theta$ as a function of $h\nu$ numerically, using the actual energy bands[11] and the polar width $\Delta\Theta = 2°$ FWHM of our analyzer. For the majority band, we obtain, e.g., $\Delta E = 0.6$ and 5.2 meV for $h\nu = 6.5$ and 11 eV, respectively. These values are only about 0.1 times the mean free path contributions (35) at the corresponding photon energies. We calculate this small additional broadening separately for each photon energy, incorporating it into the model approximately by adding the corresponding value to the mean-free-path contribution (35).

We investigate the form of the resolution function of our photoelectron spectrometer using the experimental Pt photoelectron spectrum around E_F. Since the polycrystalline Pt sample is at $T = 12$ K, the Fermi edge can be regarded as a step function and the negative derivative of the experimental Pt spectrum is the spectrometer function: It turns out to be Gaussian to a good approximation, verifying our assumption for R_{jk} in Eq. (2). As an aside we remark that any other symmetrical spectrometer function can be used with our scheme by simply introducing it in Eq. (2). In contrast, an asymmetrical spectrometer function requires a more elaborate overhaul of various equations. We also take the energy dependence of the width parameter σ of R_{jk} approximately into account by determining σ through a least-squares fit of the Fermi-Dirac function to the measured Pt Fermi edge at each of the ten different photon energies individually. We further improve the accuracy of the convolution (1) and (2) in the neighborhood of E_F by replacing the steep Fermi-Dirac function f at E_i by its spectral weight

$$\overline{f}_i = 1/\Delta \int_{E_i - \Delta/2}^{E_i + \Delta/2} dE \; f(E), \tag{36}$$

where $\Delta = 10$ meV is the step width between neighboring data points.

We also comment on the possible influence of the Ni(111) surface state,[18] which we clearly identify in the spectra taken with p polarization by its surface sensitivity, comparing the results obtained from a clean surface with that exposed to 0.25 L of O_2: At $h\nu = 7.5$ eV, the pronounced peak at $E \approx E \approx -0.3$ eV observed for the clean surface is strongly reduced in intensity after the O_2 adsorption and the position of the peak shifts to $E \approx -0.26$ eV, which is the position of the majority spin peak for s polarization at $h\nu = 7.5$ eV in

Fig. 10. The -0.3-eV peak of the clean surface is thus likely to be a superposition of the surface state and of the majority peak. In accordance with the high degree of polarization of the synchrotron radiation and the selection rules,[19] we do not find any indication of this surface state with Λ_1 symmetry in the spectra of Fig. 10, taken with *s* polarization. As a further test, we look for the Λ_1 $3d$ bulk state, which produces a pronounced peak at $E \approx -1.7$ eV for $h\nu = 15$ eV and *p* polarization, while no trace above a smooth background can be detected for *s* polarization. Thus any residual contribution of the surface state to the spectra of Fig. 10 must be very small and can safely be neglected.

We finally mention that, although available, we do not include spectra with $h\nu > 11$ eV in the Bayesian analysis since the mean-free-path contribution (35) is getting too large compared to the Im Σ term of Eqs. (29) or (32). For the same reason we fail in our attempt to perform a similar Bayesian analysis of Pt(111) spectra since the slope of the initial band is too large in this case.

D. Least-squares fit

We perform preliminary least-squares fits at all $h\nu$ to obtain the best adjustment for the six parameters specified above. For $h\nu \geq 8.5$ eV, the exchange splitting and the intensity ratio of the two lines turn out to be $\Delta_x = E_\uparrow - E_\downarrow \approx 0.3$ eV and $I_\uparrow / I_\downarrow \approx 1$, respectively, where $I_\uparrow / I_\downarrow$ is determined by integrating the optimized lines from -10 to 10 eV prior to the truncation by the Fermi-Dirac function. We expect this Δ_x value from the raw data and from earlier results[15] and the $I_\uparrow / I_\downarrow$ value from the fact that the corresponding transitions occur at closely neighboring points of the Brillouin zone and that the states involved differ only in their spin orientation otherwise. On the other hand, we find a strong increase of $I_\uparrow / I_\downarrow$ for both models in the $h\nu < 8.5$ eV range. The highest value $I_\uparrow / I_\downarrow \approx 35$ is obtained for the MFL model at $h\nu = 7.0$ eV. This unphysical increase of $I_\uparrow / I_\downarrow$ obviously serves to mask the deficiencies of the FL and MFL models since the number of parameters is too high at these low $h\nu$ where the minority peak is centered above E_F. We therefore eliminate $C_\downarrow$ as an adjustable parameter in this region by calculating it from $C_\uparrow$ and the average intensity ratio of the spectra with $h\nu \geq 8.5$ eV, which is $I_\uparrow / I_\downarrow = 1.26$ and 0.98 for the FL and MFL cases, respectively. This is not sufficient for $h\nu < 7.5$ eV, where only a small fraction of the minority peak contributes to the measured spectra since unphysically large Δ_x would be obtained. We therefore fix Δ_x in addition for $h\nu < 7.5$ eV at the average value of the Δ_x obtained at $h\nu \geq 7.5$ eV, which are 0.324 and 0.300 eV for the FL and MFL cases, respectively. We also use $\beta_2 = 1.192\ \text{eV}^{-1}$ and $\beta_1 = 0.245$ as a global values valid for all $h\nu$, determined iteratively from the condition $\Sigma \chi^2(m) \rightarrow \min$, where the sum runs over the ten photon energies. The constant describing the background electrons contributes between about 10% and 30% of the difference between the maximum count and the count above E_F, with the lowest values obtained for the FL model and high $h\nu$.

The resulting optimized models m are presented in Fig. 11. Their convolutions $G(m)$ are compared with the measured spectra in Fig. 10 and the corresponding $\chi^2(m)$ together with their sums are listed in columns 3 and 6 of Table I. The most striking difference between the measured spectra and the $G(m)$ occurs for $h\nu$ between 7.0 and 9.0 eV, where the FL model reproduces the sharpness of the truncated minority peak close to E_F much better than the MFL model, also reflected by $\chi^2(m_{\text{FL}}) < \chi^2(m_{\text{MFL}})$: Since the values of β_2 and β_1 are mainly determined by the untruncated majority peak in this region, the MFL minority peak close to E_F is

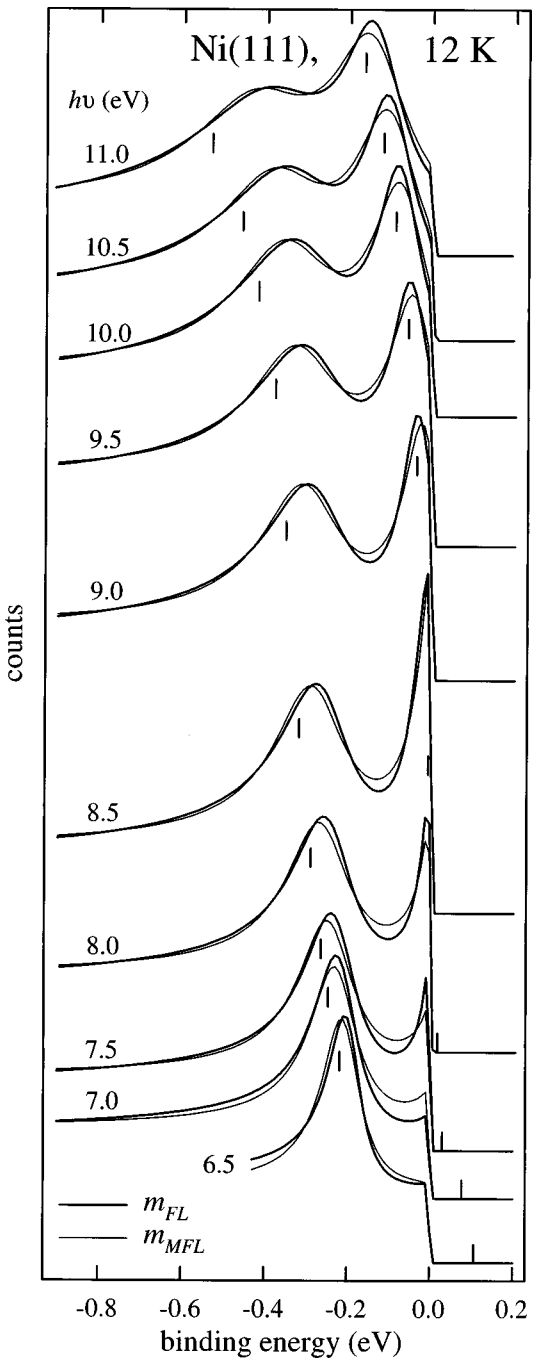

FIG. 11. Optimized FL and MFL models, shown as heavy and light solid lines, respectively. The scale conforms with the normalization of the data chosen in Fig. 10. The tick marks give the binding energies E_s for the FL peaks. Notice that the tick marks do not coincide with the peak maxima, which is one characteristic feature of the FL (and MFL) many-body spectral function. The minority FL peak is seen to cross E_F between $h\nu = 8.0$ and 8.5 eV.

TABLE I. Numerical results of the least-squares fit and of the Bayesian analysis.

$h\nu$ (eV)	n	Fermi liquid			Marginal Fermi liquid			
		$\chi^2(m)$	$\chi^2(\hat{a})$	α_0	$\chi^2(m)$	$\chi^2(\hat{a})$	α_0	$\log_{10}(r_p)$
6.5	64	155.2	60.3	0.0973	104.9	64.1	0.231	−2.04
7.0	111	304.6	105.1	0.0439	1874.2	89.4	0.0104	6.16
7.5	111	489.9	119.5	0.0340	1318.3	109.6	0.0115	5.66
8.0	111	582.3	178.2	0.0279	1950.6	152.9	0.00615	5.08
8.5	111	239.9	145.1	0.112	610.6	131.7	0.0249	5.59
9.0	111	185.9	132.8	0.191	249.0	116.1	0.0841	0.55
9.5	111	371.9	96.1	0.0303	264.9	105.3	0.0706	−3.05
10.0	111	235.9	107.5	0.0845	174.8	114.5	0.239	−3.51
10.5	111	224.7	101.5	0.0726	162.0	108.2	0.248	−3.84
11.0	111	198.3	95.2	0.153	611.5	87.8	0.0299	7.38
Σ_{χ^2}	1063	2988.6	1141.3		7320.8	1079.6		

necessarily broader than the corresponding FL peak. Thus the least-squares fit already indicates that the FL model is more likely in this energy region.

In hindsight, one might argue that the β parameters should be determined at low photon energies where the difference between the FL and MFL models is most apparent. We actually perform such a least-squares fit, obtaining $\beta_2 = 1.150\ \mathrm{eV}^{-1}$ and $\beta_1 = 0.208\ \mathrm{eV}^{-1}$ as the average of the optimized fit parameters for $h\nu = 7.0$, 7.5, and 8.0 eV, proceeding as explained above otherwise. The results are qualitatively similar, but quantitatively even more in favor of the FL model. For example, $\chi^2(m_{\mathrm{FL}}) > \chi^2(m_{\mathrm{MFL}})$ now occurs only at two instead of four photon energies and the corresponding values, which are 158.1 and 154.9 at 6.5 eV and 225.7 and 193.9 at 10.5 eV, are much closer together than the $\chi^2(m)$ values of Table I. We do not use this second fit procedure in the Bayesian analysis below, however, since it could be construed as unfairly favoring the FL model.

E. Bayesian analysis for Ni

The most important quantity that we obtain from the Bayesian analysis is of course the ratio $r_p = P_{\mathrm{FL}}/P_{\mathrm{MFL}}$ of the probabilities (27). We list $\log_{10}(r_p)$ in Table I for the ten spectra of Fig. 10, together with the corresponding $\chi^2(\hat{a})$ and α_0, and show the most probable spectral functions $\hat{a}$ in Fig. 12 for both models. As anticipated from the least-squares fit, the integrated joint probability P_{FL} is larger than P_{MFL} for $h\nu$ between 7.0 and 9.0 eV, with very high r_p values except at $h\nu = 9.0$ eV: The FL model is clearly much more probable for the low photon energies, in spite of $r_p \approx 0.01$ for $h\nu = 6.5$ eV. The spectral range is only about half as large in this case and the contribution of the minority peak negligible, which means that the energy dependence of the background that we neglected has a decisive influence.

The higher probability of the MFL model at $h\nu = 9.5$, 10.0, and 10.5 eV is followed by the extremely high r_p value in favor of the FL model at $h\nu = 11.0$ eV. This erratic behavior might just signify that it is more difficult to distinguish between the two models in this region since the fully developed minority peak no longer produces a sharp structure close to E_F and since the decrease of the escape depth of the photoelectron causes a much higher additional broadening of the lines: As for the $h\nu = 6.5$ eV case the neglected energy dependence of the background may again contribute more to the r_p values than the difference between the two models.

We evaluate the difference in the behavior at low and high photon energies by calculating the arithmetic mean of $\log_{10}(r_p)$ and its standard deviation corresponding to the geometric mean of r_p separately for the two regions. The result is $\overline{r_p} = 12\,000 \times 10^{\pm 1.5}$ for $h\nu$ between 6.5 and 8.5 eV and $\overline{r_p} = 0.32 \times 10^{\pm 2.1}$ for $h\nu$ between 9.0 and 11.0 eV: The FL model is about four orders of magnitude more probable than the MFL model in the low-photon-energy region, while the error margin is too large to allow a distinction between the two models for the high-photon-energy region.

The FL and MFL models obtained from the least-squares fit prior to the Bayesian analysis, which are shown as heavy and light solid lines in Fig. 11, differ from each other, while the corresponding most probable spectral functions $\hat{a}$ presented in Fig. 12 nearly coincide. The ripples corresponding to the a^{osc} component discussed above are obviously caused by the noise, as revealed by a close comparison with the data of Fig. 10. As expected, the amplitude of these ripples is smaller for the FL case for all photon energies for which $r_p \gg 1$. This is qualitatively equivalent to the results given in Fig. 7 (good model, small a^{osc}) and 6 (bad model, huge a^{osc}), where P_7/P_6 is about 10^{12}, i.e., about 10^8 times larger than the corresponding $\overline{r_p}$ of Ni in the low $h\nu$ region. The ratio of the regularization parameters α_0 listed in Table I turns out to be $\alpha_0^{(\mathrm{FL})}/\alpha_0^{(\mathrm{MFL})} > 1$ whenever $r_p > 1$, again in line with the results of Figs. 7 and 6, where the corresponding ratio is 18.0. As expected, Table I shows $\chi^2(\hat{a}_{\mathrm{FL}}) \approx \chi^2(\hat{a}_{\mathrm{MFL}}) \approx n$, which means that the convolutions $G(\hat{a}_{\mathrm{FL}})$ and $G(\hat{a}_{\mathrm{MFL}})$ fit the data within the noise; they are so close together that it is impossible to distinguish them in a figure, which is the reason that we present only $G(m_{\mathrm{FL}})$ and $G(m_{\mathrm{MFL}})$ in Fig. 10. Characteristically, $\chi^2(\hat{a}_{\mathrm{FL}}) > \chi^2(\hat{a}_{\mathrm{MFL}})$ whenever $r_p \gg 1$, in line with the $\chi^2(\hat{a})$ values of Figs. 6 and 7 where the corresponding difference is 12%: Since the α_0 is lower for the bad model, the Bayesian analysis works harder at fitting $G(\hat{a})$ to D, transferring a larger portion of the noise spectrum to $\hat{a}$ on the way, which is the comparatively larger a_{osc} component observed in Fig. 6 and for MFL cases in Fig. 12 for $h\nu$ between 7.0 and 8.5 eV.

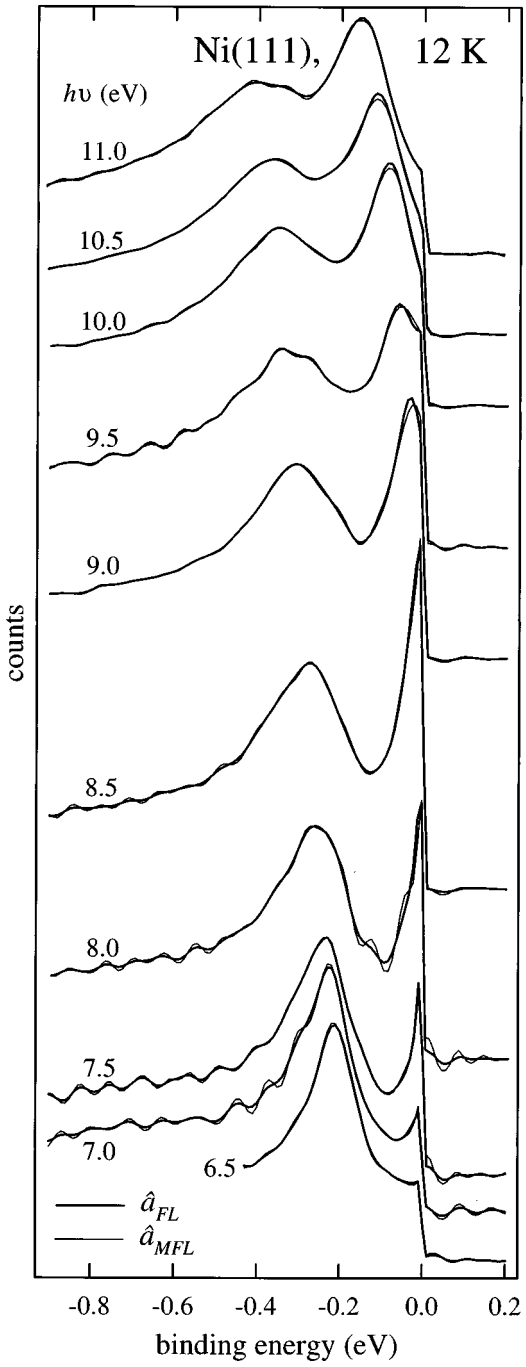

FIG. 12. The most probable spectral functions, obtained from the data of Fig. 10 and from the optimized FL and MFL models of Fig. 11, shown as heavy and light solid lines, respectively. The scale conforms with the normalization of the data.

V. SUMMARY

The continuum a of spectral functions, which, after convoluting with the spectrometer function, all fit the noisy measured spectrum D equally well, consists of the true spectral function $\tilde{a}$ containing the physics to be investigated and a continuum of unphysical oscillations a^{osc}, the convolution of which is close to zero. The iterative Bayesian deconvolution introduced in this paper is the easiest way to select from this continuum a range of spectral functions with a low $|a^{\mathrm{osc}}/\tilde{a}|$ ratio close to the most probable $\hat{a}$: Little effort is required to establish the computer code and the iteration runs fast on any personal computer. A starting model m of $\tilde{a}$ is required, with m being improved as the iteration proceeds. The unbiased model $m_u = D$ may be used, but the method also tolerates simple but smooth starting models m with a strong negative bias, e.g., $m = \mathrm{const}$, which makes it particularly useful in case no theoretical model is known. Although closely related to the Bayesian analysis, this iteration does not allow us to determine the most probable $\hat{a}$.

The Bayesian analysis is the only way to evaluate different theoretical models m in terms of the measured spectrum D quantitatively through the probability $p(D|m,\alpha_0)$ for D, given the model m, with the regularization parameter α_0 determined by $p(D|m,\alpha) \rightarrow \max$. The scientist can influence the analysis only by specifying the condition m since all results are uniquely determined by D and m. This analysis takes the continuum of all possible a into account and yields $\hat{a}$, $G(\hat{a})$ of Eq. (1), and $\chi^2(\hat{a})$ of Eq. (4) as by-products. We elucidate the role of the model by demonstrating analytically and by computer simulations that both $p(D|m,\alpha_0)$ and α_0 decrease and $|a^{\mathrm{osc}}/\tilde{a}|$ increases as m deviates from $\tilde{a}$, disqualifying the notion that one should use $m = \mathrm{const}$ to obtain an ''unbiased'' $\hat{a}$. This model, like any other model with a strong negative bias, produces $\hat{a} = \tilde{a} + a^{\mathrm{osc}}$ with a large $|a^{\mathrm{osc}}/\tilde{a}|$ ratio, with no resemblance to the original m and low values for α_0 and $p(D|m,\alpha_0)$: The Bayesian analysis is incorruptable and yields $\hat{a} \approx m \approx \tilde{a}$ with $|a^{\mathrm{osc}}/\tilde{a}| \approx 0$ for the highest $p(D|m,\alpha_0 \gg 1)$ only. Any model with $p(D|m,\alpha_0) < p(D|m_u = D,\alpha_0)$ has a negative bias and should be rejected since the $\hat{a}$ obtained from it are not trustworthy. On the other side, Ockham's razor[10] should be applied in constructing the models, using as few parameters as possible to describe the physics to be investigated and to incorporate any additional knowledge: Given two competing theoretical models containing the same small number of parameters, the Bayesian analysis yields a reliable ratio of the corresponding $p(D|m,\alpha_0)$, in spite of the imperfections of the models and the distortions and limitations of the experimental spectrum.

The parameters of the models can be optimized in principle by using the Bayesian analysis itself, but in practice the expenditure of work becomes prohibitive for more than one or two parameters. We find it equivalent but vastly more efficient to use the usual least-squares fit between D and the convolution of m for that purpose, performing the Bayesian analysis with the optimized models afterward. We also remove the severe convergence problems frequently encountered while solving the system (15) of n nonlinear equations for the n unknowns $\hat{a}_i$ by solving this system for the equivalent unknowns Δ_i defined in Eq. (7) instead. In addition, we simplify both the Bayesian iteration and the analysis considerably since we approximate the convolution by summing over the n data points only, using a normalization that minimizes the errors at the edges of the spectrum.

Applying the Bayesian analysis to the low-temperature and low-energy Ni photoemission normal to the (111) surface, the probability that these spectra are reproduced by the Fermi-liquid model turns out to be about four orders of magnitude higher than that corresponding to the marginal Fermi-liquid case for low photon energies ($6.5\ \mathrm{eV} \leqslant h\nu \leqslant 8.5\ \mathrm{eV}$) where the minority peak is strongly truncated by the Fermi-

Dirac function. This remarkable numerical result is just a consequence of the FL and MFL models treated as conditions in the Bayesian analysis, in the same sense as the numerical values of the optimized parameters contained in the models are a consequence of the least-squares fit. It is of course possible and even necessary to question the validity of these conditions, e.g., our requirement that the models cover the range from E_F to at least about $\Delta_x \approx 0.3$ eV below E_F. In the original derivation of Eq. (29), valid to arbitrary order in perturbation theory, Luttinger[20] states that this behavior is valid for E ''very near'' E_F. Is the range of our FL model too large?

We investigate this question for a simple model corresponding to the lowest order in perturbation theory, i.e., we calculate the number N of Auger processes filling the hole generated by exciting the photoelectron originally at $E(\mathbf{k}_i) = E_i$ below E_F to energies above the vacuum level. For $E(\mathbf{k}) = \hbar^2 k^2 / 2m^*$ we obtain $N \propto 3 - 5(1-\varepsilon) + 2(1-\varepsilon)^{5/2}$, where $\varepsilon = (E_F - E_i)/E_F$. Assuming $\varepsilon \ll 1$ and expanding $(1-\varepsilon)^{5/2}$, we recover $N \propto \varepsilon^2$ neglecting higher powers, i.e., Eq. (29) as expected. Evaluating E_F as the difference between the Fermi edge and the Γ_1 level, we estimate $E_F \approx 8$ eV for Ni from Fig. 9, i.e., $\varepsilon \approx \Delta_x / E_F \approx 0.04 \ll 1$ as required. Thus, in the spirit of the Fermi-liquid theory itself, our FL model assumes that the $\varepsilon \ll 1$ criterion necessary to obtain the $N \propto \varepsilon^2$ dependence for the simple model above also applies to the asymptotic behavior of the corresponding general result obtained by Luttinger.[20]

The Im Σ_2 contribution (29) is still clearly identified even around E_F since the minority peak for $h\nu = 8.5$ eV that is closest to E_F ($E_\downarrow = -7.3$ meV for FL) has the highest amplitude and the smallest effective width in both Figs. 11 and 12, in spite of the dominating escape depth contribution (35) in this region. The differences between the FL and MFL minority peaks close to E_F evident in both Figs. 11 and 10 thus also reflect the different functional form of Im Σ since the dominating contribution (35) is the same for both models. We also note that we expect only minor improvements by repeating the measurements with the same high signal-to-noise ratio but with a superior energy resolution since the escape depth contribution (35) is comparable to the half-width of the spectrometer function around $h\nu = 8.5$ eV already: Essentially, the same most probable deconvolutions $\hat{a}$ as in Fig. 12 and similar probability ratios will be obtained since these $\hat{a}$ are close to the true spectral functions $\tilde{a}$, given the optimized models and the high signal-to-noise ratio of the data.

The huge probability ratio in favor of the FL model obtained for the low photon energies where the minority peak crosses E_F is in stark contrast to the suggestion that the MFL theory might apply to the $3d$ metals[5] and it casts doubt on its validity for the high-temperature superconductors as well. At higher photon and binding energies, deviations from the simple quadratic dependence of Im Σ on E are expected, but the uncertainty in the spectral dependence of the background combined with the drastic decrease of the escape depth of the photoelectrons prevents a quantitative investigation of Im $\Sigma(E)$ in this region.

ACKNOWLEDGMENTS

V. Dose helped us start our investigation of the Bayesian analysis by sending a copy of the Skilling paper (Ref. 2) and R. Jelitto helped us get through it: We thank them both. In the final stage of preparing the manuscript, E. Mohler and W. Schwarz suggested several important improvements, which we gratefully acknowledge. This work was supported by the BMBF under the Contract No. 05 605RFA 0.

[1] P. H. van Cittert, Z. Phys. **69**, 298 (1931).
[2] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougère (Kluwer Academic, Dordrecht, 1990), pp. 341–350.
[3] W. von der Linden, M. Donath, and V. Dose, Phys. Rev. Lett. **71**, 899 (1993).
[4] W. von der Linden, Appl. Phys. A: Mater. Sci. Process. **60**, 155 (1995).
[5] A. Santoni and F. J. Himpsel, Phys. Rev. B **43**, 1305 (1991).
[6] P. M. Varma, P. B. Littlewod, S. Schmitt-Rink, E. Abrahams, and A. E. Ruckenstein, Phys. Rev. Lett. **63**, 1996 (1989).
[7] See, e.g., J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, New York, 1964).
[8] D. Marquardt, J. Soc. Ind. Appl. Math. **11**, 431 (1963).
[9] See, e.g., *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[10] A. J. M. Garrett, in *Maximum Entropy and Bayesian Methods*, edited by W. Grandy and L. Schick (Kluwer Academic, Norwell, 1991).
[11] N. V. Smith, R. Lässer, and S. Chiang, Phys. Rev. B **2**, 793 (1982).
[12] R. Claessen, R. O. Anderson, J. W. Allen, C. G. Olson, C. Janowitz, W. P. Ellis, S. Harm, M. Kalning, R. Manzke, and M. Skibowski, Phys. Rev. Lett. **69**, 808 (1992).
[13] U. Gerhardt (private communication), quoted by J. A. Knapp, F. J. Himpsel, and D. E. Eastman, Phys. Rev. B **19**, 4952 (1972).
[14] M. P. Seah and W. A. Dench, Surf. Interface Anal. **1**, 2 (1979).
[15] C. J. Maetz, U. Gerhardt, E. Dietz, A. Ziegler, and R. J. Jelitto, Phys. Rev. Lett. **48**, 1686 (1982).
[16] E. Dietz, H. Becker, and U. Gerhardt, Phys. Rev. Lett. **36**, 1397 (1976).
[17] D. Shirley, Phys. Rev. B **5**, 4709 (1972).
[18] F. J. Himpsel and D. E. Eastman, Phys. Rev. Lett. **41**, 507 (1978).
[19] J. Hermanson, Solid State Commun. **22**, 9 (1977).
[20] J. M. Luttinger, Phys. Rev. **121**, 942 (1961).